The Annals of Statistics
2013, Vol. 41, No. 1, 296–322
DOI: 10.1214/12-AOS1082


# A CRAMÉR MODERATE DEVIATION THEOREM FOR HOTELLING'S $T^2$-STATISTIC WITH APPLICATIONS TO GLOBAL TESTS

By Weidong Liu[1] and Qi-Man Shao[2]

*Shanghai Jiao Tong University and Chinese University of Hong Kong*

A Cramér moderate deviation theorem for Hotelling's $T^2$-statistic is proved under a finite $(3+\delta)$th moment. The result is applied to large scale tests on the equality of mean vectors and is shown that the number of tests can be as large as $e^{o(n^{1/3})}$ before the chi-squared distribution calibration becomes inaccurate. As an application of the moderate deviation results, a global test on the equality of $m$ mean vectors based on the maximum of Hotelling's $T^2$-statistics is developed and its asymptotic null distribution is shown to be an extreme value type I distribution. A novel intermediate approximation to the null distribution is proposed to improve the slow convergence rate of the extreme distribution approximation. Numerical studies show that the new test procedure works well even for a small sample size and performs favorably in analyzing a breast cancer dataset.

**1. Introduction.** Consider the following $m$ simultaneous tests:

$$H_{0i}:\boldsymbol{\mu}_{1i}=\boldsymbol{\mu}_{2i} \quad \text{versus} \quad H_{1i}:\boldsymbol{\mu}_{1i}\neq\boldsymbol{\mu}_{2i} \tag{1.1}$$

for $1\le i\le m$, where $\boldsymbol{\mu}_{1i}$ and $\boldsymbol{\mu}_{2i}$ are $d_i\ge 1$-dimensional mean vectors, and $d_i$ are uniformly bounded. When $d_i=1$, the multiple testing problem (1.1) has been extensively studied. A common statistical method is the two sample $t$-test together with multiple comparison procedure by controlling the familywise error rate (FWER) or the false discovery rate (FDR). The theo-

Received August 2012; revised December 2012.

[1]Supported by NSFC, Grant 11201298, the Program for Professor of Special Appointment (Eastern Scholar) at Shanghai Institutions of Higher Learning, the Foundation for the Author of National Excellent Doctoral Dissertation of PR China and the startup fund from SJTU.

[2]Supported in part by Hong Kong RGC UST603710 and CUHK2130344.

*AMS 2000 subject classifications.* Primary 62E20; secondary 62H15, 60F10.

*Key words and phrases.* Cramér moderate deviation, Hotelling's $T^2$-statistic, global tests, simultaneous hypothesis tests, FDR, brain structure, gene selection.

retical justification of this method can be found in Fan, Hall and Yao (2007). Although not much attention has been paid to the multivariate case $d_i > 1$, (1.1) has arisen from several important applications including shape analysis of brain structures and gene selection.

- *Shape analysis of brain structures.* There is a growing interest in statistical shape analysis within the neuroimaging community; see Styner et al. (2006), Zhao et al. (2008), Gerardina et al. (2009). Styner et al. (2006) developed a widely-used software to locate significant shape changes between healthy and pathological brain structures. The final and most important step in Styner et al. (2006) procedure is the simultaneous testing of (1.1) with $\boldsymbol{\mu}_{1i}$ and $\boldsymbol{\mu}_{2i}$ being mean vectors of 3 coordinates of surface points. The number of tests $m$ can be hundreds or even thousands and $d_i = 3$ for all $i$. In Styner et al. (2006), two sample Hotelling's $T^2$-statistics $T_{ni}^2$ were used for each $H_{0i}$ and Benjamini–Hochberg procedure was used to control the FDR.
- *Gene selection.* In the breast cancer dataset analyzed by Martens et al. (2005), every gene corresponds to a two to six-dimensional vector that represents the DNA methylation status of CpG sites. Dimension $d_i$ is between 2 to 6. In Martens et al. (2005), two sample Hotelling's $T^2$-statistics and Benjamini–Hochberg FDR correction were used to identify the significantly different genes between two patient groups.

It is well known that Hotelling's $T^2$-statistic is asymptotically chi-squared distributed when the underlying distribution has a finite second moment. This provides a natural way to estimate $p$-values. In the "large $m$ small $n$" statistical analysis, the true $p$-values are typically small, of order $O(1/m)$ in FDR procedure. A basic question is:

> with how many tests can the chi-squared distribution calibration be applied before the tests become inaccurate?

As discussed in Fan, Hall and Yao (2007) and Liu and Shao (2010), the question can be answered with Cramér-type moderate deviation results. The moderate deviation behavior for $t$-statistic is now well-understood, however, a Cramér type moderate deviation theorem for Hotelling's $T^2$-statistic is still not available. The main purpose of this paper is to establish the moderate deviation theorem for Hotelling's $T^2$-statistic (one-sample and two-sample). We shall prove that under a finite $(3+\delta)$th moment, Hotelling's $T^2$-statistic $T_n^2$ satisfies

$$\frac{\mathsf{P}(T_n^2 \geq x^2)}{\mathsf{P}(\chi^2(d) \geq x^2)} \to 1$$

uniformly for $x \in [0, o(n^{1/6}))$. Consequently, the number of tests can be as large as $e^{o(n^{1/3})}$ before the chi-squared distribution calibration becomes inaccurate; see (2.2).

As an application of the moderate deviation result, we consider the global testing

$$H_0 : \boldsymbol{\mu}_{1i} = \boldsymbol{\mu}_{2i} \qquad \text{for all } 1 \le i \le m \quad \text{against}$$
$$H_1 : \boldsymbol{\mu}_{1i} \ne \boldsymbol{\mu}_{2i} \qquad \text{for some } i. \tag{1.2}$$

In shape analysis of brain structures with $d_i = 3$, the global test (1.2) is often used to determine whether two brain shapes between two groups of subjects are different or not; see Cao and Worsley (1999), Taylor and Worsley (2008). In gene selection [Martens et al. (2005)], (1.2) has been used to test whether the endocrine therapy is effective on DNA methylation status. Here we are particularly interested in the alternative hypothesis that the locations where $\boldsymbol{\mu}_{1i} \ne \boldsymbol{\mu}_{2i}$ are sparse. For example, in the brain structures, the shape differences are commonly assumed to be confined to a small number of isolated regions inside the whole brain. In this paper, we shall propose a testing procedure based on the maximum of Hotelling's $T^2$-statistics. The proposed test procedure shares several advantages. It is quite robust to the tails of the underlying distribution and the dependence structure. It converges to the given significance level with a rate of $\sqrt{(\log m)^5/n}$. A numerical study shows that the test procedure works quite well even for small samples.

The rest of our paper is organized as follows. In Section 2, we state Cramér moderate deviation results for Hotelling's $T^2$-statistic. In Section 3, we introduce our test procedure for the global test (1.2). Theoretical results of the robustness on the tails and dependence structures are given. The power of the test procedure is also investigated. A numerical study is carried out in Section 4, in which we compare our test procedure to some existing test procedures. The proofs of the main results are postponed to Section 5.

## 2. A Cramér type moderate deviation theorem for Hotelling's $T^2$-statistic.

The properties of Hotelling's $T^2$-statistic under normality are well known [Anderson (2003)]. Large and moderate deviations (logarithm of the tail probabilities) were obtained in Dembo and Shao (2006). In this section, we shall establish a Cramér moderate deviation theorem for Hotelling's $T^2$-statistic. For Student $t$-statistic, the Cramér moderate deviation result was first obtained by Shao (1999) under a finite third moment and the result was extended to self-normalized sums of independent random variables in Jing, Shao and Wang (2003). We refer to de la Peña, Lai and Shao (2009) for a systematic presentation on the self-normalized limit theory and its statistical applications.

Let $\{\mathbf{X}_1, \ldots, \mathbf{X}_{n_1}\}$ and $\{\mathbf{Y}_1, \ldots, \mathbf{Y}_{n_2}\}$ be two groups of i.i.d. $d$-dimensional random vectors with mean vectors $\boldsymbol{\mu}_1$ and $\boldsymbol{\mu}_2$ and covariance matrices $\boldsymbol{\Sigma}_1$ and $\boldsymbol{\Sigma}_2$, respectively. Assume that $\{\mathbf{X}_1, \ldots, \mathbf{X}_{n_1}\}$ and $\{\mathbf{Y}_1, \ldots, \mathbf{Y}_{n_2}\}$ are

independent and $\mathbf{\Sigma}_1$ and $\mathbf{\Sigma}_2$ are positive definite. Let

$$\bar{\mathbf{X}} = \frac{1}{n_1}\sum_{k=1}^{n_1}\mathbf{X}_k, \qquad \bar{\mathbf{Y}} = \frac{1}{n_2}\sum_{k=1}^{n_2}\mathbf{Y}_k$$

be the sample means and

$$\mathbf{V}_{n1} = \frac{1}{n_1}\sum_{k=1}^{n_1}(\mathbf{X}_k - \bar{\mathbf{X}})(\mathbf{X}_k - \bar{\mathbf{X}})', \qquad \mathbf{V}_{n2} = \frac{1}{n_2}\sum_{k=1}^{n_2}(\mathbf{Y}_k - \bar{\mathbf{Y}})(\mathbf{Y}_k - \bar{\mathbf{Y}})'$$

be the sample covariance matrices, where for a vector $\mathbf{a}$, $\mathbf{a}'$ denotes its transpose. The two sample Hotelling's $T^2$-statistic is then defined by

$$T_n^2 = (\bar{\mathbf{X}} - \bar{\mathbf{Y}})'\left(\frac{1}{n_1}\mathbf{V}_{n1} + \frac{1}{n_2}\mathbf{V}_{n2}\right)^{-1}(\bar{\mathbf{X}} - \bar{\mathbf{Y}}).$$

Let $n_1 \asymp n_2$ denote the inequality $c_1 \le n_1/n_2 \le c_2$ for some positive constants $c_1$ and $c_2$. The following result gives a Cramér type moderate deviation for Hotelling's $T^2$-statistic.

THEOREM 2.1. *Suppose that $n_1 \asymp n_2$, $\mathsf{E}\|\mathbf{X}_1\|^{3+\delta} < \infty$ and $\mathsf{E}\|\mathbf{Y}_1\|^{3+\delta} < \infty$ for some $\delta > 0$. Then, under $\boldsymbol{\mu}_1 = \boldsymbol{\mu}_2$*

$$\frac{\mathsf{P}(T_n^2 \ge x^2)}{\mathsf{P}(\chi^2(d) \ge x^2)} \to 1 \qquad \text{as } n \to \infty \tag{2.1}$$

*uniformly for $x \in [0, o(n^{1/6}))$, where $n = n_1 + n_2$.*

Theorem 2.1 shows that the true distribution of $T_n^2$ can be well approximated by $\chi^2(d)$ distribution uniformly in the interval $[0, o(n^{1/3}))$ under the finite $(3+\delta)$th moment. Let $F_n(x) = \mathsf{P}(T_n^2 \ge x | \boldsymbol{\mu}_1 = \boldsymbol{\mu}_2)$ and $F(x) = \mathsf{P}(\chi^2(d) \ge x)$. Then, the true $p$-value is $p_n = F_n(T_n^2)$ and the estimated $p$-value is $\hat{p}_n = F(T_n^2)$. Thus by (2.1),

$$\left|\frac{\hat{p}_n}{p_n} - 1\right| I\{p_n \ge e^{-o(n^{1/3})}\} = o(1). \tag{2.2}$$

This provides a theoretical justification of the accuracy of the estimated $p$-values by the chi-squared distribution used in B-H FDR correction method. We refer to Fan, Hall and Yao (2007) and Liu and Shao (2010) for more detailed discussion on the relations between the Cramér type moderate deviation and the accuracy of the estimated $p$-values used in large scale tests.

For one-sample Hotelling's $T^2$-statistic, we have a similar result.

THEOREM 2.2. *Suppose that $\mathsf{E}\|\mathbf{X}_1\|^{3+\delta} < \infty$ for some $\delta > 0$. Then*

$$\frac{\mathsf{P}(n_1(\bar{\mathbf{X}} - \boldsymbol{\mu}_1)'\mathbf{V}_{n1}^{-1}(\bar{\mathbf{X}} - \boldsymbol{\mu}_1) \ge x^2)}{\mathsf{P}(\chi^2(d) \ge x^2)} \to 1 \qquad \text{as } n_1 \to \infty \tag{2.3}$$

*uniformly for $x \in [0, o(n_1^{1/6}))$.*

The proof of Theorem 2.2 is completely similar to that of Theorem 2.1 and so will be omitted.

REMARK 2.1. As proved by Shao (1999) and Jing, Shao and Wang (2003), (2.1) and (2.3) hold under finite third moments when $d = 1$ and the range $[0, o(n^{1/6}))$ is the widest possible. We conjecture that (2.1) and (2.3) remain valid for $d \geq 2$ under a finite third moment and that the range $[0, o(n^{1/6}))$ is optimal.

**3. Global testing.** In this section, we are interested in the global testing (1.2), that is,

$$H_0 : \boldsymbol{\mu}_{1i} = \boldsymbol{\mu}_{2i} \qquad \text{for all } 1 \leq i \leq m \quad \text{against}$$
$$H_1 : \boldsymbol{\mu}_{1i} \neq \boldsymbol{\mu}_{2i} \qquad \text{for some } i.$$

where $\boldsymbol{\mu}_{1i}$ and $\boldsymbol{\mu}_{2i}$ are $d_i$-dimensional mean vectors of random vectors $\mathbf{X}^i$ and $\mathbf{Y}^i$, respectively.

Write $\mathbf{a} = (\boldsymbol{\mu}'_{11}, \ldots, \boldsymbol{\mu}'_{1m})$ and $\mathbf{b} = (\boldsymbol{\mu}'_{21}, \ldots, \boldsymbol{\mu}'_{2m})$. Most of existing works on the global tests are focused on the alternative that $\mathbf{a} - \mathbf{b}$ is either sparse or dense. When the alternative is sparse, the commonly used test statistic is the maximum of univariate $t$-statistics and the higher criticism (HC$^*$) test procedure [Donoho and Jin (2004), Hall and Jin (2010)]. On the other hand, if the signals are dense, then the squared sum type test statistics have been used [Chen and Qin (2010)]. In this section, we focus on the sparse alternative hypothesis. The main difference between the current paper and the previous works is that the sparse signals appear in groups and that the underlying distributions are not necessarily normal and the components may not have an ordered structure. For the sparse case, it has been proved in Donoho and Jin (2004) that the higher criticism statistic enjoys some optimal properties with respect to the detection region. On the other hand, the independence between variables plays an important role in the control of type I errors of the higher criticism statistic. The simulation in Section 4 shows that HC$^*$ statistic may not be robust against the dependence and may fail to control the type I error. In contrast, our test procedure introduced below is robust to dependence, as shown by Theorems 3.1–3.4 and the simulation.

Suppose that we have two groups of i.i.d. observations

$$\mathcal{X} = \{\mathbf{X}_k^1, \ldots, \mathbf{X}_k^m; 1 \leq k \leq n_1\} \quad \text{and} \quad \mathcal{Y} = \{\mathbf{Y}_k^1, \ldots, \mathbf{Y}_k^m; 1 \leq k \leq n_2\}$$

with mean vectors $\{\boldsymbol{\mu}_{11}, \ldots, \boldsymbol{\mu}_{1m}\}$ and $\{\boldsymbol{\mu}_{21}, \ldots, \boldsymbol{\mu}_{2m}\}$, respectively. The two groups of observations $\mathcal{X}$ and $\mathcal{Y}$ are independent. Let $T_{ni}^2$ be the two sample Hotelling's $T^2$-statistics based on $\{\mathbf{X}_k^i; 1 \leq k \leq n_1\}$ and $\{\mathbf{Y}_k^i; 1 \leq k \leq n_2\}$. We introduce our test procedure as follows.

*Case* 1. $d_i \equiv d$. Let $\mathbf{W}_{1,k}$, $1 \le k \le n_1$, and $\mathbf{W}_{2,k}$, $1 \le k \le n_2$ be i.i.d. multivariate normal vectors with mean zero and covariance matrix $\mathbf{I}_d$. Let

$$F_{n_1,n_2}(y) = \mathsf{P}(T_n^{*2} \ge y), \tag{3.1}$$

where $T_n^{*2}$ is the two sample Hotelling's $T^2$-test statistic based on $\{\mathbf{W}_{1,k}\}$ and $\{\mathbf{W}_{2,k}\}$. For given $0 < \alpha < 1$, let $y_n(\alpha)$ satisfy

$$\exp(-mF_{n_1,n_2}(y_n(\alpha))) = 1 - \alpha. \tag{3.2}$$

Note that $1 - F_{n_1,n_2}(y)$ is closely related to $F$ distribution. In general, we can use simulation to obtain $y_n(\alpha)$. Our test procedure for (1.2) is $\Phi_\alpha^*$, where

$$\Phi_\alpha^* = I\Big\{\max_{1\le i\le m} T_{ni}^2 \ge y_n(\alpha)\Big\}. \tag{3.3}$$

The hypothesis $H_0$ is rejected whenever $\Phi_\alpha^* = 1$.

*Case* 2. *$d_i$ may be different*. Let $F_{n_1,n_2,d_i}(y)$ be defined as in (3.1) with $d$ being replaced with $d_i$. Let $G_{n_1,n_2,d_i}(y) = 1 - F_{n_1,n_2,d_i}(y)$. We now define

$$\Phi_\alpha^\dagger = I\Big\{\max_{1\le i\le m} G_{n_1,n_2,d_i}(T_{ni}^2) \ge g_m(\alpha)\Big\}$$

with $g_m(\alpha) = 1 + m^{-1}\log(1-\alpha)$. The hypothesis $H_0$ is rejected whenever $\Phi_\alpha^\dagger = 1$. Note that $\Phi_\alpha^\dagger = \Phi_\alpha^*$ if $d_i \equiv d$.

REMARK 3.1. By Theorem 3.1, $\max_{1\le i\le m} T_{ni}^2$ converges to the extreme I type distribution. It seems natural to define the following test $\Phi_\alpha$:

$$\Phi_\alpha = I\Big\{\max_{1\le i\le m} T_{ni}^2 \ge 2\log m + (d-2)\log\log m + q_\alpha\Big\}, \tag{3.4}$$

where $q_\alpha = -2\log(\Gamma(d/2)) - 2\log\log(1-\alpha)^{-1}$. The hypothesis $H_0$ is rejected whenever $\Phi_\alpha = 1$. However, it is well known that the rate of convergence to the extreme distribution is very slow [see Liu, Lin and Shao (2008)]. On the other hand, the intermediate approximation given in Theorem 3.3 can substantially improve the convergence rate. This leads to our test procedure $\Phi_\alpha^*$. Numerical results in Section 4 show that $\Phi_\alpha^*$ outperforms $\Phi_\alpha$ significantly and it works well even when the sample size is small.

3.1. *The limiting distribution of* $\max_{1\le i\le m} T_{ni}^2$. In this subsection, we show that the type I error of $\Phi_\alpha^*$ will converges to $\alpha$ under some mild moment conditions and dependence structure. To this end, we need to establish the limiting distribution of $\max_{1\le i\le m} T_{ni}^2$ under $H_0$. Let $\mathbf{\Sigma}_i = \mathbf{\Sigma}_{i1} + \frac{n_1}{n_2}\mathbf{\Sigma}_{i2}$, where $\mathbf{\Sigma}_{i1}$ and $\mathbf{\Sigma}_{i2}$ are the covariance matrices of $\mathbf{X}^i$ and $\mathbf{Y}^i$, respectively. Define

$$\mathbf{\Gamma}_{ij} = \mathbf{\Sigma}_i^{-1/2}\bigg(\mathsf{Cov}(\mathbf{X}^i, \mathbf{X}^j) + \frac{n_1}{n_2}\mathsf{Cov}(\mathbf{Y}^i, \mathbf{Y}^j)\bigg)\mathbf{\Sigma}_j^{-1/2}.$$

The matrix $\mathbf{\Gamma}_{ij}$ characterizes the dependence structure between $\{\mathbf{X}^i, \mathbf{Y}^i\}$ and $\{\mathbf{X}^j, \mathbf{Y}^j\}$. For example, when $n_1 = n_2$ and $\mathbf{\Sigma}_{i1} = \mathbf{\Sigma}_{i2}$,

$$\mathbf{\Gamma}_{ij} = \tfrac{1}{2}\,\mathsf{Cov}(\mathbf{\Sigma}_{i1}^{-1/2}\mathbf{X}^i, \mathbf{\Sigma}_{j1}^{-1/2}\mathbf{X}^j) + \tfrac{1}{2}\,\mathsf{Cov}(\mathbf{\Sigma}_{i2}^{-1/2}\mathbf{Y}^i, \mathbf{\Sigma}_{j2}^{-1/2}\mathbf{Y}^j)$$

is the sum of two matrices. When $d = 1$ and $\mathbf{\Sigma}_{i1} = \mathbf{\Sigma}_{i2}$, then $\mathbf{\Gamma}_{ij} = \rho_{ij1}$, where $\rho_{ij1}$ is the correlation coefficient between $\mathbf{X}^i$ and $\mathbf{X}^j$. For $0 < r < 1$, let

$$\Lambda(r) = \{1 \le i \le m : \|\mathbf{\Gamma}_{ij}\| \ge r \text{ for some } j \ne i\},$$

where $\|\cdot\|$ is the spectral norm. $\Lambda(r)$ is a subset of $\{1, 2, \dots, m\}$ in which $\{\mathbf{X}^i, \mathbf{Y}^i\}$ can be highly correlated with other random vectors. Let $\mathbf{R}_1 = (r_{ij1})$ and $\mathbf{R}_2 = (r_{ij2})$ be the correlation matrices of the random vectors $((\mathbf{X}^1)', \dots, (\mathbf{X}^m)')$ and $((\mathbf{Y}^1)', \dots, (\mathbf{Y}^m)')$, respectively. For some $\gamma > 0$, let

$$s_j(m) = \mathrm{Card}\{1 \le i \le m : |r_{ij1}| \ge (\log m)^{-1-\gamma} \text{ or } |r_{ij2}| \ge (\log m)^{-1-\gamma}\}.$$

We need the following condition on the dependence structure.

(C1) Suppose that $\mathrm{Card}(\Lambda(r)) = o(m)$ for some $0 < r < 1$ and

$$\max_{1 \le j \le p} s_j(m) = O(m^\rho)$$

for all $\rho > 0$. Assume that $\min_{1 \le i \le p}\{\lambda_{\min}(\mathbf{\Sigma}_i)\} \ge \tau$ for some $\tau > 0$, where $\lambda_{\min}(\mathbf{\Sigma}_i)$ is the smallest eigenvalue of $\mathbf{\Sigma}_i$.

The dependence condition (C1) is mild. In (C1), $o(m)$ vectors $\{\mathbf{X}^i, \mathbf{Y}^i\}$, $i \in \Lambda(r)$, can be highly correlated with other random vectors. Every $\{\mathbf{X}^i, \mathbf{Y}^i\}$ can be highly correlated with $s_i(m)$ vectors and weakly correlated with the remaining vectors. The dependence in (C1) is more general than "clumpy dependence" [Storey and Tibshirani (2001)] and may be a more realistic form of dependence in DNA microarrays. See also Hall and Wang (2010) who noted that short-range dependence, and more specially, $k$-dependence structure, are often observed in DNA microarrays.

The next condition is on the moment of the underlying distributions and the relation between the sample sizes and dimension $m$. We assume that $m$ is a function of $n = n_1 + n_2$ and $m \to \infty$ as $n \to \infty$.

(C2) Suppose that $\max_{1 \le i \le m} \mathsf{E}(\|\mathbf{X}^i\|^{3+\delta} + \|\mathbf{Y}^i\|^{3+\delta}) \le \kappa$ for some $\kappa > 0$ and $\delta > 0$, $n_1 \asymp n_2$ and $\log m = o(n^{1/3})$.

THEOREM 3.1. *Under $H_0$, $d_i \equiv d$,* (C1) *and* (C2), *we have as $n \to \infty$,*

$$\begin{aligned} &\mathsf{P}\Big(\max_{1 \le i \le m} T_{ni}^2 - 2\log m + (2-d)\log\log m \le y\Big) \\ &\qquad \to \exp\left(-\frac{1}{\Gamma(d/2)} e^{-y/2}\right) \end{aligned} \tag{3.5}$$

*for any $y \in R$.*

It follows from Theorem 2.1 that

$$y_n(\alpha) = 2\log m + (d-2)\log\log m + q_\alpha + o(1),$$

which together with Theorem 3.1, yields the following theorem.

THEOREM 3.2. *Under $H_0$, $d_i \equiv d$, (C1) and (C2), we have as $n \to \infty$,*

$$\mathsf{P}(\Phi^*_\alpha = 1) \to \alpha. \tag{3.6}$$

REMARK 3.2. When $d_i$ are different, we have a similar result as Theorem 3.2. Under $H_0$, (C1) and (C2), we have as $n \to \infty$,

$$\mathsf{P}(\Phi^\dagger_\alpha = 1) \to \alpha \tag{3.7}$$

for any $0 < \alpha < 1$. The proof of (3.7) is similar to that of Theorem 3.1 and hence will be omitted.

As mentioned earlier, the convergence rate of (3.5) is very slow. In testing diagonal covariance matrix problem, Liu, Lin and Shao (2008) proposed to use an intermediate approximation and proved that the rate of convergence can be of order of $\sqrt{(\log m)^5/n}$. Here we give a similar intermediate approximation to the distribution of $\max_{1\le i\le m} T^2_{ni}$.

Let $\Theta_j$ be the set of indices such that $T^2_{nj}$ is independent with $(T^2_{ni}; i \in \Theta_j)$ and put $s_j(m) = m - \mathrm{Card}(\Theta_j)$.

(C1$^*$) Suppose that $\mathrm{Card}(\Lambda(r)) = O(m^\xi)$ for some $0 < r < 1$ and $0 \le \xi < 1$. Assume that $\max_{1\le j\le m} s_j(m) = O(m^\rho)$ for some $0 < \rho < (1-r)/(1+r)$.

(C2$^*$) Suppose that $\max_{1\le i\le m} \mathsf{E}(\|\mathbf{X}^i\|^{3+\delta} + \|\mathbf{Y}^i\|^{3+\delta}) \le \kappa$ for some $\kappa > 0$ and $\delta > 0$, $c_1 \le n_1/n_2 \le c_2$ for some $c_1 > 0$ and $c_2 > 0$ and $\log m = o(n^{1/3})$.

(C3$^*$) Suppose that $\mathbf{\Sigma}_{1i} = \mathbf{\Sigma}_{2i}$ for $1 \le i \le m$. We assume that $\mathbf{X}^i$ and $\mathbf{Y}^i$ can be written as the transforms of independent components:

$$\mathbf{X}^i = \mathbf{\Sigma}^{1/2}_{1i}\mathbf{Z}_{1i} + \boldsymbol{\mu}_{1i} \quad \text{and} \quad \mathbf{Y}^i = \mathbf{\Sigma}^{1/2}_{2i}\mathbf{Z}_{2i} + \boldsymbol{\mu}_{2i},$$

where $\mathsf{E}\mathbf{Z}_{1i} = 0$, $\mathsf{Cov}(\mathbf{Z}_{1i}) = \mathbf{I}$ and $\mathsf{E}\mathbf{Z}_{2i} = 0$, $\mathsf{Cov}(\mathbf{Z}_{2i}) = \mathbf{I}$ and the components in $\mathbf{Z}_{1i}$ and $\mathbf{Z}_{2i}$ are independent.

(C1$^*$) is a technical condition. It allows $T^2_{nj}$ be dependent with $O(m^\rho)$ others. By (C1$^*$), we can use the Poisson approximation in Arratia, Goldstein and Gordon (1989). (C3$^*$) is also required for technical reason. It can be avoided if we assume that $\max_{1\le i\le m} \mathsf{E}e^{t(\|\mathbf{X}^i_1\| + \|\mathbf{Y}^i_1\|)} \le \kappa$ for some $t > 0$.

THEOREM 3.3. *Under $H_0$, $d_i \equiv d$, (C1$^*$)–(C3$^*$), we have for any $\epsilon > 0$*

$$\begin{aligned}\sup_{y\in R}&\Bigl|\mathsf{P}\Bigl(\max_{1\le i\le m} T^2_{ni} < y\Bigr) - \exp(-mF_{n_1,n_2}(y))\Bigr| \\ &\le C\left(\sqrt{\frac{(\log m)^5}{n}} + m^{\rho-(1-r)/(1+r)+\epsilon} + m^{\xi-1}\log m\right),\end{aligned} \tag{3.8}$$

*where $F_{n_1,n_2}(y)$ is defined in (3.1) and $C$ is a finite constant depending only on $\xi, r, \rho, \delta, \kappa, \epsilon, c_1, c_2$ and $d$.*

If $m \geq c_1 n^b$ for all $b > 0$, then the error rate in Theorem 3.3 is of order $\sqrt{(\log m)^5/n}$. By Theorem 3.3, we can get the following result.

THEOREM 3.4. *Under $H_0$, $d_i \equiv d$, (C1*)–(C3*), we have for any $\epsilon > 0$,*

$$\sup_{0 \leq \alpha \leq 1} |\mathsf{P}(\Phi_\alpha^* = 1) - \alpha| \leq C\left(\sqrt{\frac{(\log m)^5}{n}} + m^{\rho-(1-r)/(1+r)+\epsilon} + m^{\xi-1}\log m\right),$$

*where $C$ is given in (3.8).*

3.2. *Power result for $\Phi_\alpha^*$.* Here we consider the power of the test $\Phi_\alpha^*$.

THEOREM 3.5. *Suppose that*

$$\max_{1 \leq i \leq m} \|\boldsymbol{\Sigma}_i^{-1/2}(\boldsymbol{\mu}_{1i} - \boldsymbol{\mu}_{2i})\| \geq \sqrt{\frac{(2+\epsilon)\log m}{n_1}}$$

*for some $\epsilon > 0$. Then under* (C1) *and* (C2),

$$\mathsf{P}(\Phi_\alpha^* = 1) \to 1 \qquad \text{as } n \to \infty.$$

Theorem 3.5 shows that, in order to reject the null hypothesis correctly, we only require $\max_{1 \leq i \leq m} \|\boldsymbol{\Sigma}_i^{-1/2}(\boldsymbol{\mu}_{1i} - \boldsymbol{\mu}_{2i})\| \geq \sqrt{\frac{(2+\epsilon)\log m}{n_1}}$. The optimality of this lower bound when $d = 1$ can be found in Cai, Liu and Xia (2012). We believe this lower bound remains optimal for $d \geq 2$ under some regularity conditions.

## 4. Numerical results.

4.1. *Simulation.* In this section, we examine the numerical performance of the proposed tests $\Phi_\alpha^*$ with $d = 3$. We first compare $\Phi_\alpha^*$ with $\Phi_\alpha$ to see the improvement of the intermediate approximation and then compare $\Phi_\alpha^*$ to the higher criticism (HC*) test procedure [Donoho and Jin (2004), Hall and Jin (2010)], the test procedure proposed by Chen and Qin (2010) (C-Q) and the univariate $t$-test procedure based on $\max_{1 \leq i \leq dm} t_i^2$ (U-T), where $t_i$ is the two sample $t$-statistic based on the $i$th coordinates of the observations. The higher criticism test statistic is defined as Hall and Jin (2010)

$$\mathrm{HC}^* = \max_{j:1/q \leq p_{(j)} \leq 1/2} \left\{\frac{\sqrt{q}(j/q - p_{(j)})}{\sqrt{p_{(j)}(1-p_{(j)})}}\right\},$$

where $q = 3m$, $p_j = \mathsf{P}(|N(0,1)| \geq |t_i|)$ and $p_{(j)}$ is the $j$th $p$-value after sorting in ascending order. There are also other versions of HC* statistics [Donoho

and Jin (2004)]. They perform similarly in our numerical studies. The critical values $\alpha_n$ with significance level 0.05 are taken to be the solutions to $\mathsf{P}(\mathrm{HC}^* \geq \alpha_n) = 0.05$ under that $p_j$, $1 \leq j \leq 3m$, are i.i.d. uniform $(0,1)$ distributed random variables.

Let

$$((\mathbf{X}^1)', \ldots, (\mathbf{X}^m)') = (Z_1^1, \ldots, Z_1^{3m}) \times \mathbf{\Sigma}^{1/2},$$
$$((\mathbf{Y}^1)', \ldots, (\mathbf{Y}^m)') = (Z_2^1, \ldots, Z_2^{3m}) \times \mathbf{\Sigma}^{1/2}$$

be $3m$-dimensional random vectors with covariance matrix $\mathbf{\Sigma}$, where $\{Z_i^j\}$ are i.i.d. random variables. We consider four distributions of $Z_i^j$, $N(0,1)$, $t(5)$, exponential distribution with parameter 1 (Exp(1)), and Gamma distribution with shape and scale parameters $(2,2)$ (Gamma(2,2)). The covariance matrix $\mathbf{\Sigma}$ is taken to be:

(1) $\mathbf{\Sigma}_1 = (0.9^{|j-i|})$;
(2) $\mathbf{\Sigma}_2 = (\sigma_{ij})$, where $\sigma_{ij} = \max\{1 - |j-i|/(0.1 * (3m)), 0\}$;
(3) $\mathbf{\Sigma}_3 = (\sigma_{ij})$, where $\sigma_{ij} = \max\{1 - |j-i|/(0.8 * (3m)), 0\}$.

$\mathbf{\Sigma}_1$ is an approximately bandable matrix. $\mathbf{\Sigma}_2$ is a $0.3m$ sparse matrix which has $0.3m$ nonzero entries in each row. In $\mathbf{\Sigma}_3$, the number of nonzero entries in each row is $2.4m$ and the dependence between the variables becomes stronger than that in $\mathbf{\Sigma}_2$.

The sample sizes $(n_1, n_2)$ are taken to be $(6,12)$, $(12,24)$, $(24,48)$ and $m$ takes values $50, 100, 200, 400$. We carry out 5000 simulations to obtain the empirical sizes with nominal significance level 0.05. The results for $\mathbf{\Sigma} = \mathbf{\Sigma}_1$ are summarized in Table 1. The simulation results when $\mathbf{\Sigma}$ takes the other covariance matrices are stated in the supplement material [Liu and Shao (2013)] due to limit of space. We can see that the empirical sizes of $\Phi_\alpha^*$ and Chen and Qin's test are close to 0.05. $\Phi_\alpha^*$ still performs well when the dependence becomes stronger ($\mathbf{\Sigma} = \mathbf{\Sigma}_2$ and $\mathbf{\Sigma}_3$). However, the empirical sizes of $\Phi_\alpha$ suffer very serious distortions. This indicates the intermediate approximation in Section 3 gains a lot of improvement on the accuracy of controlling type I errors. The test procedure $\Phi_\alpha^*$ is robust to the tails of distributions and the dependence. On the other hand, the empirical sizes of $\mathrm{HC}^*$ are much larger than 0.05. This shows that $\mathrm{HC}^*$ statistic may be not robust to the dependence. We have also done additional simulations and found that, when the variables are independent but not normally distributed, $\mathrm{HC}^*$ statistic may suffer serious distortions from the nominal significance level.

To evaluate the power, we consider both approximately sparse model and dense model. Let $\boldsymbol{\mu}_{1i} = 0$ for $1 \leq i \leq m$. Set $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_{3m}) = \mathsf{E}((\mathbf{Y}^1)', \ldots, (\mathbf{Y}^m)')$ and $\sigma^2 = \mathsf{Var}(Z_1^1)$. Consider

*Model* 1 (*approximately sparse case*). Let $\mu_i = (-0.2)^{i-1} \times 2\sqrt{\sigma^2 \log m / n_2}$ for $1 \leq i \leq 3m$.

Table 1
*Comparison of empirical sizes with nominal significance level 0.05 ($\mathbf{\Sigma} = \mathbf{\Sigma}_1$)*

| | | $N(0,1)$ | | | $t(5)$ | | |
|---|---|---|---|---|---|---|---|
| | $m \setminus (n_1, n_2)$ | **(6, 12)** | **(12, 24)** | **(24, 48)** | **(6, 12)** | **(12, 24)** | **(24, 48)** |
| 50 | $\Phi_\alpha^*$ | **0.0516** | **0.0466** | **0.0430** | **0.0412** | **0.0374** | **0.0404** |
| | $\Phi_\alpha$ | 0.8965 | 0.4760 | 0.2285 | 0.8641 | 0.4312 | 0.2078 |
| | HC* | 0.5986 | 0.4348 | 0.3514 | 0.6028 | 0.4438 | 0.3534 |
| | C-Q | 0.0634 | 0.0644 | 0.0632 | 0.0646 | 0.0660 | 0.0644 |
| 100 | $\Phi_\alpha^*$ | **0.0558** | **0.0483** | **0.0508** | **0.0423** | **0.0360** | **0.0442** |
| | $\Phi_\alpha$ | 0.9694 | 0.5799 | 0.2711 | 0.9542 | 0.5315 | 0.2364 |
| | HC* | 0.7584 | 0.5228 | 0.4260 | 0.7460 | 0.5334 | 0.4100 |
| | C-Q | 0.0606 | 0.0620 | 0.0626 | 0.0642 | 0.0614 | 0.0592 |
| 200 | $\Phi_\alpha^*$ | **0.0602** | **0.0584** | **0.0515** | **0.0464** | **0.0393** | **0.0420** |
| | $\Phi_\alpha$ | 0.9958 | 0.7045 | 0.3238 | 0.9916 | 0.6380 | 0.2783 |
| | HC* | 0.9072 | 0.6492 | 0.4920 | 0.8986 | 0.6438 | 0.4672 |
| | C-Q | 0.0624 | 0.0584 | 0.0600 | 0.0566 | 0.0570 | 0.0574 |
| 400 | $\Phi_\alpha^*$ | **0.0636** | **0.0609** | **0.0495** | **0.0464** | **0.0402** | **0.0406** |
| | $\Phi_\alpha$ | 1.0000 | 0.8198 | 0.3781 | 0.9996 | 0.7571 | 0.3253 |
| | HC* | 0.9840 | 0.7876 | 0.5660 | 0.9814 | 0.7820 | 0.5642 |
| | C-Q | 0.0552 | 0.0592 | 0.0604 | 0.0508 | 0.0580 | 0.0588 |
| | | **Exp(1)** | | | **Gamma(2, 2)** | | |
| 50 | $\Phi_\alpha^*$ | **0.0355** | **0.0392** | **0.0450** | **0.0403** | **0.0468** | **0.0451** |
| | $\Phi_\alpha$ | 0.8441 | 0.4294 | 0.2226 | 0.8675 | 0.4473 | 0.2291 |
| | HC* | 0.5950 | 0.4492 | 0.3584 | 0.5924 | 0.4370 | 0.3604 |
| | C-Q | 0.0628 | 0.0622 | 0.0688 | 0.0580 | 0.0728 | 0.0666 |
| 100 | $\Phi_\alpha^*$ | **0.0404** | **0.0372** | **0.0519** | **0.0436** | **0.0414** | **0.0524** |
| | $\Phi_\alpha$ | 0.9409 | 0.5230 | 0.2625 | 0.9557 | 0.5521 | 0.2725 |
| | HC* | 0.7502 | 0.5296 | 0.4188 | 0.7640 | 0.5352 | 0.4212 |
| | C-Q | 0.0620 | 0.0626 | 0.0644 | 0.0664 | 0.0582 | 0.0598 |
| 200 | $\Phi_\alpha^*$ | **0.0408** | **0.0364** | **0.0498** | **0.0481** | **0.0435** | **0.0551** |
| | $\Phi_\alpha$ | 0.9882 | 0.6355 | 0.3105 | 0.9923 | 0.6671 | 0.3196 |
| | HC* | 0.8910 | 0.6358 | 0.4806 | 0.9042 | 0.6538 | 0.5014 |
| | C-Q | 0.0602 | 0.0608 | 0.0630 | 0.0570 | 0.0556 | 0.0610 |
| 400 | $\Phi_\alpha^*$ | **0.0460** | **0.0355** | **0.0517** | **0.0478** | **0.0449** | **0.0529** |
| | $\Phi_\alpha$ | 0.9987 | 0.7430 | 0.3671 | 0.9997 | 0.7810 | 0.3693 |
| | HC* | 0.9766 | 0.7788 | 0.5768 | 0.9838 | 0.7916 | 0.5762 |
| | C-Q | 0.0570 | 0.0590 | 0.0568 | 0.0518 | 0.0544 | 0.0572 |

*Model* 2 (*dense case*). Let $\mu_i = 0.2(-1)^{i-1} \times 2\sqrt{\sigma^2 \log m/n_2}$ for $1 \leq i \leq 3m$.

Because of the serious distortion of empirical sizes of $\Phi_\alpha$ and HC*, we do not consider the power of $\Phi_\alpha$ and HC*. We only report the power results for the normal distributions due to the high similarity of the results with other distributions. The reject region for $\max_{1\leq i\leq dm} t_i^2$ is $[y_n(\alpha), \infty)$ with $d = 1$ in

TABLE 2
*Comparison of empirical powers ($\boldsymbol{\Sigma} = \boldsymbol{\Sigma}_1$)*

| | | **Model 1** | | | **Model 2** | | |
|---|---|---|---|---|---|---|---|
| | $\boldsymbol{m \setminus (n_1, n_2)}$ | $\mathbf{(6,12)}$ | $\mathbf{(12,24)}$ | $\mathbf{(24,48)}$ | $\mathbf{(6,12)}$ | $\mathbf{(12,24)}$ | $\mathbf{(24,48)}$ |
| 50 | $\Phi_\alpha^*$ | **0.7343** | **0.9327** | **0.9758** | **0.9453** | **0.9959** | **0.9994** |
| | C-Q | 0.0755 | 0.0739 | 0.0755 | 0.1369 | 0.1343 | 0.1404 |
| | U-T | 0.0766 | 0.0938 | 0.1064 | 0.0901 | 0.0890 | 0.0862 |
| 100 | $\Phi_\alpha^*$ | **0.7489** | **0.9538** | **0.9880** | **0.9943** | **1.0000** | **1.0000** |
| | C-Q | 0.0704 | 0.0733 | 0.0720 | 0.2201 | 0.2250 | 0.2295 |
| | U-T | 0.0713 | 0.1001 | 0.0921 | 0.1019 | 0.1137 | 0.0875 |
| 200 | $\Phi_\alpha^*$ | **0.7451** | **0.9635** | **0.9937** | **0.9998** | **1.0000** | **1.0000** |
| | C-Q | 0.0761 | 0.0665 | 0.0705 | 0.4289 | 0.4365 | 0.4303 |
| | U-T | 0.0719 | 0.1058 | 0.0945 | 0.1278 | 0.1507 | 0.1160 |
| 400 | $\Phi_\alpha^*$ | **0.7520** | **0.9696** | **0.9957** | **1.000** | **1.0000** | **1.0000** |
| | C-Q | 0.0633 | 0.0634 | 0.0636 | 0.7701 | 0.7997 | 0.8007 |
| | U-T | 0.0703 | 0.1089 | 0.0951 | 0.1414 | 0.2062 | 0.1467 |

$F_{n_1,n_2}(y)$ and $y_n(\alpha)$ satisfying

$$\exp(-3mF_{n_1,n_2}(y_n(\alpha))) = 1 - \alpha.$$

This gives a much more accurate approximation than the extreme distribution (results will not be reported here).

In Table 2, we only state the results when $\boldsymbol{\Sigma} = \boldsymbol{\Sigma}_1$. The other simulation results are given in the supplement material [Liu and Shao (2013)]. Note that in model 1, $n\|\boldsymbol{\mu}\|^2/m^{1/2} \to 0$. The power of Chen and Qin (2010) is low, as shown in Table 2. The power of $\max_{1\le i\le dm} t_i^2$ is also quite low. Our test statistics $\Phi_\alpha^*$ has the highest powers which are close to one for $(n_1, n_2) = (12, 24)$ and $(24, 48)$. In the dense case model 2, our test statistics still has the highest power. We should remark that no method can uniformly outperform others over all models and there may exist certain situations where Chen and Qin's (2010) test statistic may outperform ours.

4.2. *Real data analysis.* We apply the test procedure in Section 3 to test whether the tamoxifen therapy is effective on the promoter DNA methylation status of 117 genes. The dataset consists of 123 patients, who showed the extreme types of response to tamoxifen treatment; they either had an objective response (CR + PR, 45 patients) or a progressive disease right from the start of treatment (PD, 78 patients). There are 117 genes and each gene corresponds to a 2–6-dimensional vector that represents DNA methylation status of CpG sites analyzed using a microarray-based DNA methylation detection assay. Martens et al. (2005) used the Benjamini–Hochberg (B-H) FDR procedure with the target FDR of 25% to identify genes whose promoter DNA methylation status was associated with the clinical benefit of tamoxifen therapy. Before using B-H FDR procedure, it is interesting to test

whether the tamoxifen therapy is effective on the promoter DNA methylation status of those genes.

For each gene, we calculate the Hotelling's $T^2$-statistic $T_{ni}^2$. The given significance level is $\alpha = 0.05$. The value of $\max_{1 \le i \le m} G_{n_1,n_2,d_i}(T_{ni}^2)$ is 1.0000 which is larger than $1 + m^{-1}\log(0.95) = 0.9996$. Thus, we can accept at the 0.05 significance level that the tamoxifen therapy has an effect on the promoter DNA methylation status. We found three genes, PSAT1, STMN1 and SFN, whose values of $G_{n_1,n_2,d_i}(T_{ni}^2)$ are larger than 0.9996. These three genes were also identified by Martens et al. (2005) who used B-H FDR correction and the $\chi^2$ distributions.

**5. Proof of main results.**

5.1. *Proof of Theorem 2.1.* Without loss of generality, we assume that $\boldsymbol{\mu}_1 = \boldsymbol{\mu}_2 = 0$. Since $T_n^2$ converges to a chi-squared distribution with $d$ degrees of freedom, we have for any $M > 0$

$$\lim_{n\to\infty} \sup_{0\le x\le M} \left| \frac{\mathsf{P}(T_n^2 \ge x^2)}{\mathsf{P}(\chi^2(d) \ge x^2)} - 1 \right| = 0.$$

Thus, there exists a sequence $a_n \to \infty$ such that

$$\lim_{n\to\infty} \sup_{0\le x\le a_n} \left| \frac{\mathsf{P}(T_n^2 \ge x^2)}{\mathsf{P}(\chi^2(d) \ge x^2)} - 1 \right| = 0. \tag{5.1}$$

Let $\boldsymbol{\Sigma} = \boldsymbol{\Sigma}_1 + \frac{n_1}{n_2}\boldsymbol{\Sigma}_2$ and

$$\mathbf{Z}_k = \begin{cases} \boldsymbol{\Sigma}^{-1/2}\mathbf{X}_k, & 1 \le k \le n_1, \\ -\dfrac{n_1}{n_2}\boldsymbol{\Sigma}^{-1/2}\mathbf{Y}_{k-n_1}, & n_1 + 1 \le k \le n_1 + n_2. \end{cases}$$

By the identity

$$\mathbf{x}'\mathbf{A}^{-1}\mathbf{x} = \max_{\|\theta\|=1} \frac{(\mathbf{x}'\theta)^2}{\theta'\mathbf{A}\theta}$$

for any $d \times d$ positive definite matrix $\mathbf{A}$, where $\theta$ is a $d$-dimensional vector, we have

$$\begin{aligned}\{T_n^2 \ge x^2\} = \Bigg\{ \exists\theta, \text{ s.t. } \|\theta\| = 1, &\left|\sum_{k=1}^n \theta'\mathbf{Z}_k\right| \\ &\ge x\sqrt{\sum_{k=1}^n (\theta'\mathbf{Z}_k)^2 - n_1(\theta'\bar{\mathbf{Z}}_1)^2 - n_2(\theta'\bar{\mathbf{Z}}_2)^2}\Bigg\},\end{aligned}$$

where $n = n_1 + n_2$, $\bar{\mathbf{Z}}_1 = \frac{1}{n_1}\sum_{k=1}^{n_1}\mathbf{Z}_k$ and $\bar{\mathbf{Z}}_2 = \frac{1}{n_2}\sum_{k=n_1+1}^{n}\mathbf{Z}_k$. Theorem 2.1 follows if we can prove that

$$\frac{\mathsf{P}(\exists\theta, \text{ s.t. } \|\theta\| = 1, |\sum_{k\in H}\theta'\mathbf{Z}_k| \ge x\sqrt{\sum_{k\in H}(\theta'\mathbf{Z}_k)^2})}{\mathsf{P}(\chi^2(d) \ge x^2)} \to 1 \tag{5.2}$$

uniformly for $x \in [a_n, o(n^{1/6}))$, $H = \{1, 2, \ldots, n\}$, $\{1, 2, \ldots, n_1\}$ and $\{n_1 + 1, \ldots, n\}$. In fact, (5.2) implies that, for $i = 1, 2$,

$$\frac{\mathsf{P}(\exists\theta, \text{ s.t. } \|\theta\| = 1, |\theta'\bar{\mathbf{Z}}_i| \geq 2n_i^{-1}x\sqrt{\sum_{k=1}^{n}(\theta'\mathbf{Z}_k)^2})}{\mathsf{P}(\chi^2(d) \geq 4x^2)} \to 1$$

uniformly for $x \in [a_n, o(n^{1/6}))$. Observe that

$$\begin{aligned}
&\mathsf{P}(T_n^2 \geq x^2)\\
&\quad \leq \mathsf{P}\left(\exists\theta, \text{ s.t. } \|\theta\| = 1, |\theta'\bar{\mathbf{Z}}_1| \geq 2n_1^{-1}x\sqrt{\sum_{k=1}^{n_1}(\theta'\mathbf{Z}_k)^2}\right)\\
&\qquad + \mathsf{P}\left(\exists\theta, \text{ s.t. } \|\theta\| = 1, |\theta'\bar{\mathbf{Z}}_2| \geq 2n_2^{-1}x\sqrt{\sum_{k=n_1+1}^{n}(\theta'\mathbf{Z}_k)^2}\right)\\
&\qquad + \mathsf{P}\left(\exists\theta, \text{ s.t. } \|\theta\| = 1, \frac{|\sum_{k=1}^{n}\theta'\mathbf{Z}_k|}{(\sum_{k=1}^{n}(\theta'\mathbf{Z}_k)^2)^{1/2}} \geq x(1 - 4x^2n_1^{-1} - 4x^2n_2^{-1})^{1/2}\right)\\
&\quad = (2 + o(1))\mathsf{P}(\chi^2(d) \geq 4x^2)\\
&\qquad + \mathsf{P}\left(\exists\theta, \text{ s.t. } \|\theta\| = 1, \frac{|\sum_{k=1}^{n}\theta'\mathbf{Z}_k|}{(\sum_{k=1}^{n}(\theta'\mathbf{Z}_k)^2)^{1/2}} \geq x(1 - 4x^2n_1^{-1} - 4x^2n_2^{-1})^{1/2}\right)\\
&\quad = o(1)\mathsf{P}(\chi^2(d) \geq x^2)\\
&\qquad + \mathsf{P}\left(\exists\theta, \text{ s.t. } \|\theta\| = 1, \frac{|\sum_{k=1}^{n}\theta'\mathbf{Z}_k|}{(\sum_{k=1}^{n}(\theta'\mathbf{Z}_k)^2)^{1/2}} \geq x(1 - 4x^2n_1^{-1} - 4x^2n_2^{-1})^{1/2}\right)
\end{aligned}$$

uniformly in $x \in [a_n, o(n^{1/6}))$. Similarly, we can obtain a lower bound for $\mathsf{P}(T_n^2 \geq x^2)$, which together with (5.1) and (5.2) yields (2.1).

We only prove (5.2) with $H = \{1, 2, \ldots, n\}$. The proof for the other two cases is similar. Let $3/(3 + \delta) < \beta < 1$, $\hat{\mathbf{Z}}_k = \mathbf{Z}_k I\{\|\mathbf{Z}_k\| \leq (\sqrt{n}/x)^\beta\}$ and set

$$S_n(\theta) = \sum_{k=1}^{n}\theta'\mathbf{Z}_k, \qquad S_n^{\{\mathbf{N}\}}(\theta) = \sum_{k=1, k\notin\mathbf{N}}^{n}\theta'\mathbf{Z}_k,$$

$$\hat{S}_n(\theta) = \sum_{k=1}^{n}\theta'\hat{\mathbf{Z}}_k, \qquad \hat{S}_n^{\{\mathbf{N}\}}(\theta) = \sum_{k=1, k\notin\mathbf{N}}^{n}\theta'\hat{\mathbf{Z}}_k,$$

$$\mathbf{V}_n(\theta) = \sum_{k=1}^{n}(\theta'\mathbf{Z}_k)^2, \qquad \mathbf{V}_n^{\{\mathbf{N}\}}(\theta) = \sum_{k=1, k\notin\mathbf{N}}^{n}(\theta'\mathbf{Z}_k)^2,$$

$$\hat{\mathbf{V}}_n(\theta) = \sum_{k=1}^{n}(\theta'\hat{\mathbf{Z}}_k)^2, \qquad \hat{\mathbf{V}}_n^{\{\mathbf{N}\}}(\theta) = \sum_{k=1, k\notin\mathbf{N}}^{n}(\theta'\hat{\mathbf{Z}}_k)^2,$$

where $\mathbf{N}$ is an index set. By the fact that [see (5.7) in Jing, Shao and Wang (2003)]

$$\{s+t\ge x\sqrt{c+t^2}\}\subset\{s\ge(x^2-1)^{1/2}\sqrt{c}\} \tag{5.3}$$

for any $s,t\in R$, $c\ge 0$ and $x\ge 1$, we have

$$\begin{aligned}
&\mathsf{P}(\exists\theta,\text{ s.t. }\|\theta\|=1,|S_n(\theta)|\ge x\sqrt{\mathbf{V}_n(\theta)})\\
&\quad\le \mathsf{P}(\exists\theta,\text{ s.t. }\|\theta\|=1,|\hat{S}_n(\theta)|\ge x\sqrt{\hat{\mathbf{V}}_n(\theta)})\\
&\qquad+\sum_{j=1}^n \mathsf{P}(\exists\theta,\text{ s.t. }\|\theta\|=1,|S_n^{\{j\}}(\theta)|\ge\sqrt{x^2-1}\sqrt{\mathbf{V}_n^{\{j\}}(\theta)},A_j) \qquad (5.4)\\
&\quad= \mathsf{P}(\exists\theta,\text{ s.t. }\|\theta\|=1,|\hat{S}_n(\theta)|\ge x\sqrt{\hat{\mathbf{V}}_n(\theta)})\\
&\qquad+\sum_{j=1}^n \mathsf{P}(\exists\theta,\text{ s.t. }\|\theta\|=1,|S_n^{\{j\}}(\theta)|\ge\sqrt{x^2-1}\sqrt{\mathbf{V}_n^{\{j\}}(\theta)})P(A_j),
\end{aligned}$$

where

$$A_j=\{\|\mathbf{Z}_j\|\ge(\sqrt{n}/x)^\beta\}\qquad\text{for } 1\le j\le n.$$

Repeating (5.4) and inequality (5.3) $m$ times, we get

$$\begin{aligned}
&\mathsf{P}(\exists\theta,\text{ s.t. }\|\theta\|=1,|S_n(\theta)|\ge x\sqrt{\mathbf{V}_n(\theta)})\\
&\quad\le \mathsf{P}(\exists\theta,\text{ s.t. }\|\theta\|=1,|\hat{S}_n(\theta)|\ge x\sqrt{\hat{\mathbf{V}}_n(\theta)})+\sum_{l=1}^m\hat{U}_l+U_{m+1},
\end{aligned}$$

where

$$\begin{aligned}
\hat{U}_l=&\sum_{j_1=1}^n\cdots\sum_{j_l=1}^n\left[\prod_{k=1}^l\mathsf{P}(A_{j_k})\right]\\
&\times\mathsf{P}(\exists\theta,\text{ s.t. }\|\theta\|=1,|\hat{S}_n^{\{j_1,\dots,j_l\}}(\theta)|\ge\sqrt{x^2-l}\sqrt{\hat{\mathbf{V}}_n^{\{j_1,\dots,j_l\}}(\theta)})
\end{aligned}$$

and

$$U_{m+1}=\sum_{j_1=1}^n\cdots\sum_{j_{m+1}=1}^n\prod_{k=1}^{m+1}\mathsf{P}(A_{j_k}).$$

Let $m=[x^2/2]$ for $x\ge 4$. We have

$$\begin{aligned}
U_{m+1}&=\left(\sum_{k=1}^n\mathsf{P}(\|\mathbf{Z}_k\|\ge(\sqrt{n}/x)^\beta)\right)^{m+1}\\
&\le e^{-m\log q_n}=o(1)\mathsf{P}(\chi^2(d)\ge x),
\end{aligned} \tag{5.5}$$

where

$$q_n = (n(x/\sqrt{n})^{\beta(3+\delta)}\mathsf{E}(\|\mathbf{X}_1\|^{3+\delta} + \|\mathbf{Y}_1\|^{3+\delta}))^{-1} \to \infty.$$

The proof of (5.2) now relies on the Cramér-type moderate theorem for self-normalized truncated variables given below.

PROPOSITION 5.1. *Assume that* $\mathrm{Card}(\mathbf{N}) = O(x^2)$. *Then we have*

$$\begin{aligned} \mathsf{P}(\exists\theta,\ s.t.\ \|\theta\| = 1, |\hat{S}_n^{\{\mathbf{N}\}}(\theta)| \geq x\sqrt{\hat{\mathbf{V}}_n^{\{\mathbf{N}\}}(\theta)}) \\ = (1 + o(1))\mathsf{P}(\chi^2(d) \geq x^2) \end{aligned} \tag{5.6}$$

*uniformly in* $x \in [a_n, o(n^{1/6}))$.

The proof of Proposition 5.1 will be given in the next subsection. Let us now finish the proof of (5.2).

Using the same arguments as in the proof of inequality (5.5) and by Proposition 5.1, we have

$$\begin{aligned} \sum_{l=1}^{m} \hat{U}_l &\leq C\sum_{l=1}^{m} \mathsf{P}(\chi^2(d) \geq x^2 - l)\exp(-l\log q_n) \\ &= o(1)\mathsf{P}(\chi^2(d) \geq x^2) \end{aligned}$$

uniformly in $x \in [a_n, o(n^{1/6}))$. Hence,

$$\mathsf{P}(\exists\theta, \text{ s.t. } \|\theta\| = 1, |S_n(\theta)| \geq x\sqrt{\mathbf{V}_n(\theta)}) \leq (1 + o(1))\mathsf{P}(\chi^2(d) \geq x^2)$$

uniformly in $x \in [a_n, o(n^{1/6}))$. To establish the lower bound, we note that

$$\begin{aligned} &\mathsf{P}(\exists\theta, \text{ s.t. } \|\theta\| = 1, |S_n(\theta)| \geq x\sqrt{\mathbf{V}_n(\theta)}) \\ &\quad \geq \mathsf{P}(\exists\theta, \text{ s.t. } \|\theta\| = 1, |\hat{S}_n(\theta)| \geq x\sqrt{\hat{\mathbf{V}}_n(\theta)}) \\ &\qquad - \sum_{j=1}^{n} \mathsf{P}(\exists\theta, \text{ s.t. } \|\theta\| = 1, |\hat{S}_n^{\{j\}}(\theta)| \geq \sqrt{x^2 - 1}\sqrt{\hat{\mathbf{V}}_n^{\{j\}}(\theta)})P(A_j). \end{aligned}$$

It follows from Proposition 5.1 again that

$$\mathsf{P}(\exists\theta, \text{ s.t. } \|\theta\| = 1, |S_n(\theta)| \geq x\sqrt{\mathbf{V}_n(\theta)}) \geq (1 + o(1))\mathsf{P}(\chi^2(d) \geq x^2)$$

uniformly in $x \in [a_n, o(n^{1/6}))$. This completes the proof of (5.2) and hence Theorem 2.1.

5.2. *Proof of Proposition 5.1.* We start with the Cramér type moderate deviation theorem for non-self-normalized sum.

LEMMA 5.1. *Let* $\mathrm{Card}(\mathbf{N}) = O(x^2)$. *We have*

$$\mathsf{P}(\exists\theta,\ s.t.\ \|\theta\| = 1, |\hat{S}_n^{\{\mathbf{N}\}}(\theta)| \geq x\sqrt{n_1}) = (1 + o(1))\mathsf{P}(\chi^2(d) \geq x^2)$$

*uniformly in* $x \in [4, o(n^{1/6}))$.

To prove Lemma 5.1, we need the following lemma by Lin and Liu (2009). The definition $|\cdot|_d$ below is a slightly different from that in Lin and Liu (2009), but the proof is exactly the same.

LEMMA 5.2. *Let $\xi_{n,1},\dots,\xi_{n,k_n}$ be independent random vectors with mean zero and values in $R^d$, and $S_n=\sum_{i=1}^{k_n}\xi_{n,i}$. Assume that $\|\xi_{n,i}\|\le c_nB_n^{1/2}$, $1\le i\le k_n$, for some $c_n\to 0$, $B_n\to\infty$ and*

$$\|B_n^{-1}\mathsf{Cov}(\xi_{n,1}+\cdots+\xi_{n,k_n})-I_d\|\le C_0c_n^2,$$

*where $I_d$ is a $d\times d$ identity matrix and $C_0$ is a positive constant. Suppose that $\beta_n:=B_n^{-3/2}\sum_{i=1}^{k_n}\mathsf{E}\|\xi_{n,i}\|^3\to 0$. Then for all $n\ge n_0$ ($n_0$ is given below)*

$$\begin{aligned}&|\mathsf{P}(|S_n|_d\ge x)-\mathsf{P}(|N|_d\ge x/B_n^{1/2})|\\&\quad\le o(1)\mathsf{P}(|N|_d\ge x/B_n^{1/2})\\&\qquad+C_d\left(\exp\left(-\frac{\delta_n^2\min(c_n^{-2},\beta_n^{-2/3})}{8d}\right)+\exp\left(\frac{C_dc_n^2}{\beta_n^2\log\beta_n}\right)\right),\end{aligned}$$

*uniformly for $x\in[B_n^{1/2},\delta_n\min(c_n^{-1},\beta_n^{-1/3})B_n^{1/2}]$, with any $\delta_n\to 0$ and $\delta_n\min(c_n^{-1},\beta_n^{-1/3})\to\infty$, where $N$ is a centered normal random vector with covariance matrix $I_d$; $|\cdot|_d$ denotes $|\mathbf{z}|_d=\min\{\|\mathbf{x}_i\|:1\le i\le d/q\}$, $\mathbf{z}=(\mathbf{x}_1,\dots,\mathbf{x}_{d/q})$, $\mathbf{x}_i\in R^q$ and $d/q$ is an integer; $o(1)$ is bounded by $A_n:=A(\delta_n+\beta_n)$, $A$ is a positive constant depending only on $d$;*

$$n_0=\min\{n:\forall k\ge n,c_k^2\le C_{01},\delta_k\le C_{02},\beta_k\le C_{03}\},$$

*where $C_{01}$, $C_{02}$ and $C_{03}$ are some positive constants depending only on $d$ and $C_0$.*

PROOF OF LEMMA 5.1. Let $\xi_{nk}=\hat{\mathbf{Z}}_k-\mathsf{E}\hat{\mathbf{Z}}_k$, $B_n=n_1$ and $c_n=2n_1^{-1/2}(\sqrt{n}/x)^{\beta}$ in Lemma 5.2. By the inequalities $\beta>3/(3+\delta)$ and $x=o(n^{1/6})$,

$$\begin{aligned}\left\|B_n^{-1}\mathsf{Cov}\left(\sum_{k=1}^n\xi_{nk}\right)-I_d\right\|&\le C\max_{1\le k\le n}\mathsf{E}\|\mathbf{Z}_k\|^2I\{\|Z_k\|\ge(\sqrt{n}/x)^{\beta}\}\\&\le C(x/\sqrt{n})^{(1+\delta)\beta}\le Cc_n^2.\end{aligned}$$

By letting $\delta_n\to 0$ sufficiently slow, we have

$$\exp\left(-\frac{\delta_n^2\min(c_n^{-2},\beta_n^{-2/3})}{8d}\right)+\exp\left(\frac{C_dc_n^2}{\beta_n^2\log\beta_n}\right)=o(1)\mathsf{P}(\chi^2(d)\ge x^2)$$

uniformly in $x\in[4,o(n^{1/6}))$. This proves Lemma 5.1. $\square$

PROOF OF PROPOSITION 5.1. Observe that

$$
\begin{aligned}
&\mathsf{P}(\exists\theta, \text{ s.t. } \|\theta\|=1, |\hat{S}_n^{\{\mathbf{N}\}}(\theta)| \geq x\sqrt{\hat{\mathbf{V}}_n^{\{\mathbf{N}\}}(\theta)})\\
&\quad \leq \mathsf{P}(\exists\theta, \text{ s.t. } \|\theta\|=1, |\hat{S}_n^{\{\mathbf{N}\}}(\theta)| \geq x\sqrt{n_1(1-\varepsilon_n x^{-2})})\\
&\qquad + \mathsf{P}(\exists\theta, \text{ s.t. } \|\theta\|=1, |\hat{S}_n^{\{\mathbf{N}\}}(\theta)| \geq x\sqrt{\hat{\mathbf{V}}_n^{\{\mathbf{N}\}}(\theta)}, E_n(\theta))
\end{aligned}
$$

and

$$
\begin{aligned}
&\mathsf{P}(\exists\theta, \text{ s.t. } \|\theta\|=1, |\hat{S}_n^{\{j\}}(\theta)| \geq x\sqrt{\hat{\mathbf{V}}_n^{\{j\}}(\theta)})\\
&\quad \geq \mathsf{P}(\exists\theta, \text{ s.t. } \|\theta\|=1, |\hat{S}_n^{\{j\}}(\theta)| \geq x\sqrt{n_1(1+\varepsilon_n x^{-2})})\\
&\qquad - \mathsf{P}(\exists\theta, \text{ s.t. } \|\theta\|=1, |\hat{S}_n^{\{j\}}(\theta)| \geq x\sqrt{n_1(1+\varepsilon_n x^{-2})}, F_n(\theta)),
\end{aligned}
$$

where $\varepsilon_n \to 0$ which will be specified later and

$$
\begin{aligned}
E_n(\theta) &= \{\hat{\mathbf{V}}_n^{\{\mathbf{N}\}}(\theta) \leq n_1(1-\varepsilon_n x^{-2})\},\\
F_n(\theta) &= \{\hat{\mathbf{V}}_n^{\{j\}}(\theta) \geq n_1(1+\varepsilon_n x^{-2})\}.
\end{aligned}
$$

Also note that

$$
\mathsf{P}(\exists\theta, \text{ s.t. } \|\theta\|=1, |\hat{S}_n^{\{\mathbf{N}\}}(\theta)| \geq x\sqrt{n_1}) = \mathsf{P}(|\hat{S}_n^{\{\mathbf{N}\}}|_d \geq x\sqrt{n_1})
$$

with $q=d$. By Lemma 5.1, we have

$$
\mathsf{P}(\exists\theta, \text{ s.t. } \|\theta\|=1, |\hat{S}_n^{\{\mathbf{N}\}}(\theta)| \geq x\sqrt{n_1(1\pm\varepsilon_n x^{-2})}) = (1+o(1))\mathsf{P}(\chi^2(d) \geq x^2)
$$

uniformly in $x \in [a_n, o(n^{1/6}))$. So it suffices to prove the following lemma. □

LEMMA 5.3. *Let* $\text{Card}(\mathbf{N}) = O(x^2)$. *We have*

$$
\begin{aligned}
&\mathsf{P}(\exists\theta, \textit{ s.t. } \|\theta\|=1, |\hat{S}_n^{\{\mathbf{N}\}}(\theta)| \geq x\sqrt{\hat{\mathbf{V}}_n^{\{\mathbf{N}\}}(\theta)}, E_n(\theta))\\
&\quad = o(1)\mathsf{P}(\chi^2(d) \geq x^2)
\end{aligned}
\tag{5.7}
$$

*and*

$$
\begin{aligned}
&\mathsf{P}(\exists\theta, \textit{ s.t. } \|\theta\|=1, |\hat{S}_n^{\{j\}}(\theta)| \geq x\sqrt{n_1(1+\varepsilon_n x^{-2})}, F_n(\theta))\\
&\quad = o(1)\mathsf{P}(\chi^2(d) \geq x^2)
\end{aligned}
\tag{5.8}
$$

*uniformly in* $x \in [a_n, o(n^{1/6}))$.

PROOF. We only prove (5.7) because the proof of (5.8) is similar. Let $b = x/\sqrt{n_1}$. Then for $0 < \varepsilon_n < 1/2$,

$$\begin{aligned}
&\{\hat{S}_n^{\{\mathbf{N}\}}(\theta) \ge x\sqrt{\hat{\mathbf{V}}_n^{\{\mathbf{N}\}}(\theta)}, E_n(\theta)\} \\
&\quad \subset \{2b\hat{S}_n^{\{\mathbf{N}\}}(\theta) - b^2\hat{\mathbf{V}}_n^{\{\mathbf{N}\}}(\theta) \ge x^2 - \varepsilon_n^2, E_n(\theta)\} \\
&\qquad \cup \{\hat{S}_n^{\{\mathbf{N}\}}(\theta) \ge x\sqrt{\hat{\mathbf{V}}_n^{\{\mathbf{N}\}}(\theta)}, 2xb\sqrt{\hat{\mathbf{V}}_n^{\{\mathbf{N}\}}(\theta)} < b^2\hat{\mathbf{V}}_n^{\{\mathbf{N}\}}(\theta) + x^2 - \varepsilon_n^2, E_n(\theta)\}.
\end{aligned}$$

We can choose $n_d$ points $\theta_j$, $1 \le j \le n_d$, with $\|\theta_j\| = 1$ and $n_d \le n^{2d}$, such that for any $\|\theta\| = 1$, $\|\theta - \theta_j\| \le Cn^{-2}$ for some $1 \le j \le n_d$. So we have

$$\begin{aligned}
&\mathsf{P}\biggl(\bigcup_{\|\theta\|=1}\{2b\hat{S}_n^{\{\mathbf{N}\}}(\theta) - b^2\hat{\mathbf{V}}_n^{\{\mathbf{N}\}}(\theta) \ge x^2 - \varepsilon_n^2, E_n(\theta)\}\biggr) \\
&\quad \le \sum_{j=1}^{n_d}\mathsf{P}(2b\hat{S}_n^{\{\mathbf{N}\}}(\theta_j) - b^2\hat{\mathbf{V}}_n^{\{\mathbf{N}\}}(\theta_j) \ge x^2 - \varepsilon_n^2 - n_1^{-1}, \\
&\qquad\qquad\qquad \mathbf{V}_n^{\{\mathbf{N}\}}(\theta_j) \le n_1(1 - \varepsilon_n x^{-2}) + n_1^{-1}) \\
&\quad \le \sum_{j=1}^{n_d}\mathsf{P}(2b\hat{S}_n^{\{\mathbf{N}\}}(\theta_j) - b^2(\hat{\mathbf{V}}_n^{\{\mathbf{N}\}}(\theta_j) - \mathsf{E}\hat{\mathbf{V}}_n^{\{\mathbf{N}\}}(\theta_j)) \\
&\qquad\qquad + t(\mathsf{E}\hat{\mathbf{V}}_n^{\{\mathbf{N}\}}(\theta_j) - \hat{\mathbf{V}}_n^{\{\mathbf{N}\}}(\theta_j)) \\
&\qquad\qquad \ge 2x^2 - \varepsilon_n^2 - n_1^{-1} - O(nb^3) + tn_1\varepsilon_n x^{-2} - O(ntb)) \\
&\quad =: \sum_{j=1}^{n_d} I_j.
\end{aligned}$$

Let $t = (x/\sqrt{n})^{2-\gamma}$ with $0 < \gamma < \beta(1+\delta) - 1$ and $\max\{(x^2/n)^{\gamma/4}, a_n^{-1/2}\} \le \varepsilon_n \to 0$. We use Corollary 5 of Sakhanenko (1991) to bound $I_j$. Let

$$\xi_k = 2b\theta_j'\hat{\mathbf{Z}}_k - 2b\mathsf{E}\theta_j'\hat{\mathbf{Z}}_k - (b^2 - t)((\theta_j'\hat{\mathbf{Z}}_k)^2 - \mathsf{E}(\theta_j'\hat{\mathbf{Z}}_k)^2), \qquad k \notin \mathbf{N}.$$

Then $|\xi_k| = O(1)$, $B_n^2 = \sum_{k\notin\mathbf{N}} \mathsf{E}\xi_k^2 = 4x^2 + O(1)nb^3$, and for any bounded $h$,

$$L(h) = \sum_{k\notin\mathbf{N}} \mathsf{E}|\xi_k|^3 \max\{e^{h\xi_k}, 1\} = O(1)nb^3,$$

where $O(1)$ are bounded by some absolute constants. Let

$$y_n(x) = 2x^2 - \varepsilon_n^2 - n_1^{-1} - O(nb^3) + tn_1\varepsilon_n x^{-2} - O(ntb).$$

By Corollary 5 of Sakhanenko (1991) and direct calculations, we obtain that

$$\begin{aligned}
I_j &= (1 - \Phi(y_n(x)/B_n))(1 + O(x^3/\sqrt{n})) \\
&= O(1)x^{-1}\exp(-x^2/2 - (n/x^2)^{\gamma/2})
\end{aligned}$$

uniformly in $x \in [a_n, o(n^{1/6}))$. Hence, it follows that

$$\begin{aligned} &\mathsf{P}\biggl(\bigcup_{\|\theta\|=1}\{2b\hat{S}_n^{\{\mathbf{N}\}}(\theta) - b^2\hat{\mathbf{V}}_n^{\{\mathbf{N}\}}(\theta) \geq x^2 - \varepsilon_n^2, E_n(\theta)\}\biggr) \\ &\qquad = o(1)\mathsf{P}(\chi^2(d) \geq x^2) \end{aligned} \tag{5.9}$$

uniformly in $x \in [a_n, o(n^{1/6}))$.

Observe that

$$\begin{aligned} &\{\hat{S}_n^{\{\mathbf{N}\}}(\theta) \geq x\sqrt{\hat{\mathbf{V}}_n^{\{\mathbf{N}\}}(\theta)}, 2xb\sqrt{\hat{\mathbf{V}}_n^{\{\mathbf{N}\}}(\theta)} < b^2\hat{\mathbf{V}}_n^{\{\mathbf{N}\}}(\theta) + x^2 - \varepsilon_n^2, E_n(\theta)\} \\ &\quad \subset \{\hat{S}_n^{\{\mathbf{N}\}}(\theta) \geq x\sqrt{\hat{\mathbf{V}}_n^{\{\mathbf{N}\}}(\theta)}, b^2\hat{\mathbf{V}}_n^{\{\mathbf{N}\}}(\theta) > x^2 + \varepsilon_n x, E_n(\theta)\} \\ &\qquad \cup \{\hat{S}_n^{\{\mathbf{N}\}}(\theta) \geq x\sqrt{\hat{\mathbf{V}}_n^{\{\mathbf{N}\}}(\theta)}, b^2\hat{\mathbf{V}}_n^{\{\mathbf{N}\}}(\theta) < x^2 - \varepsilon_n x, E_n(\theta)\}. \end{aligned} \tag{5.10}$$

By Lemma 5.1,

$$\begin{aligned} &\mathsf{P}\biggl(\bigcup_{\|\theta\|=1}\{\hat{S}_n^{\{\mathbf{N}\}}(\theta) \geq x\sqrt{\hat{\mathbf{V}}_n^{\{\mathbf{N}\}}(\theta)}, b^2\hat{\mathbf{V}}_n^{\{\mathbf{N}\}}(\theta) > x^2 + \varepsilon_n x, E_n(\theta)\}\biggr) \\ &\qquad \leq \mathsf{P}\biggl(\bigcup_{\|\theta\|=1}\{\hat{S}_n^{\{\mathbf{N}\}}(\theta) \geq \sqrt{(x^2 + \varepsilon_n x)n_1}\}\biggr) \\ &\qquad = (1 + o(1))\mathsf{P}(\chi^2(d) \geq x^2 + \varepsilon_n x) \\ &\qquad = o(1)\mathsf{P}(\chi^2(d) \geq x^2) \end{aligned}$$

uniformly in $[a_n, o(n^{1/6}))$ for any $a_n \to \infty$. For the second term on the right-hand side of (5.10),

$$\begin{aligned} &\mathsf{P}\biggl(\bigcup_{\|\theta\|=1}\{\hat{S}_n^{\{\mathbf{N}\}}(\theta) \geq x\sqrt{\hat{\mathbf{V}}_n^{\{\mathbf{N}\}}(\theta)}, b^2\hat{\mathbf{V}}_n^{\{\mathbf{N}\}}(\theta) < x^2 - \varepsilon_n x, E_n(\theta)\}\biggr) \\ &\qquad \leq \sum_{k=1}^{[x]}\mathsf{P}\biggl(\bigcup_{\|\theta\|=1}\{\hat{S}_n^{\{\mathbf{N}\}}(\theta) \geq x\sqrt{\hat{\mathbf{V}}_n^{\{\mathbf{N}\}}(\theta)}, \\ &\qquad\qquad \hat{\mathbf{V}}_n^{\{\mathbf{N}\}}(\theta) \in [n_1(1 - \varepsilon_n(k+1)/x), n_1(1 - \varepsilon_n k/x)]\}\biggr) \\ &\qquad\quad + \mathsf{P}\biggl(\bigcup_{\|\theta\|=1}\{\hat{\mathbf{V}}_n^{\{\mathbf{N}\}}(\theta) \leq n_1(1 - \varepsilon_n/2)\}\biggr). \end{aligned} \tag{5.11}$$

For the last term above, we use the Bernstein inequality and obtain

$$\mathsf{P}\biggl(\bigcup_{\|\theta\|=1}\{\hat{\mathbf{V}}_n^{\{\mathbf{N}\}}(\theta) \leq n_1(1 - \varepsilon_n/2)\}\biggr)$$

$$\begin{aligned}
&\le \sum_{j=1}^{n_d} \mathsf{P}(\hat{\mathbf{V}}_n^{\{\mathbf{N}\}}(\theta_j) \le n_1(1-\varepsilon_n/2)+n^{-1}) \\
&\le \sum_{j=1}^{n_d} \mathsf{P}(\mathsf{E}\hat{\mathbf{V}}_n^{\{\mathbf{N}\}}(\theta_j) - \hat{\mathbf{V}}_n^{\{\mathbf{N}\}}(\theta_j) \ge n_1(\varepsilon_n/2 + O(x/\sqrt{n}))) \\
&\le \exp\biggl(-\frac{n_1(\varepsilon_n/2+O(x/\sqrt{n}))^2}{2b^{-2\beta}+4b^{-2\beta}(\varepsilon_n/2+O(x/\sqrt{n}))/3}\biggr) \\
&= o(1)\mathsf{P}(\chi^2(d)\ge x^2)
\end{aligned}$$

uniformly in $[a_n, o(n^{1/6}))$. For the first term in (5.11), as in the proof of (5.9) using Corollary 5 of Sakhanenko (1991), we can show that

$$\begin{aligned}
&\mathsf{P}\biggl(\bigcup_{\|\theta\|=1}\{\hat{S}_n^{\{\mathbf{N}\}}(\theta) \ge x\sqrt{\hat{\mathbf{V}}_n^{\{\mathbf{N}\}}(\theta)}, \\
&\qquad \hat{\mathbf{V}}_n^{\{\mathbf{N}\}}(\theta) \in [n_1(1-\varepsilon_n(k+1)/x), n_1(1-\varepsilon_n k/x)]\}\biggr) \\
&\quad \le \mathsf{P}\biggl(\bigcup_{\|\theta\|=1}\{\hat{S}_n^{\{\mathbf{N}\}}(\theta) \ge x\sqrt{n_1(1-\varepsilon_n(k+1)/x)}, \\
&\qquad\qquad \hat{\mathbf{V}}_n^{\{\mathbf{N}\}}(\theta) \le n_1(1-\varepsilon_n k/x)\}\biggr) \\
&\quad \le \mathsf{P}\biggl(\bigcup_{\|\theta\|=1}\{b\hat{S}_n^{\{\mathbf{N}\}}(\theta) + t(\mathsf{E}\hat{\mathbf{V}}_n^{\{\mathbf{N}\}}(\theta) - \hat{\mathbf{V}}_n^{\{\mathbf{N}\}}(\theta)) \\
&\qquad\qquad \ge x\sqrt{n_1(1-\varepsilon_n(k+1)/x)} + n_1 t\varepsilon_n k/x + O(ntb)\}\biggr) \\
&\quad \le Cn_d x^{-1}\exp(-x^2/2 - c_0 x^{-\gamma} n^{\gamma/2}\varepsilon_n) \\
&\quad = o(1)\mathsf{P}(\chi^2(d)\ge x^2)
\end{aligned}$$

uniformly in $[a_n, o(n^{1/6}))$. This completes the proof of Lemma 5.3. □

5.3. *Proof of Theorem 3.1.* Let $x_n = (2\log m + (d-2)\log\log m + x)^{1/2}$. Note that by Theorem 2.1,

$$\mathsf{P}\Bigl(\max_{i\in\Lambda(r)} T_{ni}^2 \ge x_n^2\Bigr) \le C\,\mathrm{Card}(\Lambda(r))m^{-1} = o(1).$$

It suffices to prove that

$$\mathsf{P}\Bigl(\max_{i\notin\Lambda)(r)} T_{ni}^2 \ge x_n^2\Bigr) \to \exp\biggl(-\frac{1}{\Gamma(d/2)}\exp(-x/2)\biggr).$$

Since $\mathrm{Card}(\Lambda(r)) = o(m)$, without loss of generality, we can assume that $\Lambda(r) = \varnothing$, that is, $\max_{1\le i<j\le m}\|\Gamma_{ij}\| \le r$ for some $r < 1$. Otherwise, we only need to replace $\max_{1\le i\le m}(\cdot)$ below by $\max_{1\le i\le m, i\notin\Lambda(r)}(\cdot)$ and the proof remains the same. As in the proof of Theorem 2.1, we set

$$\mathbf{Z}_k^i = \begin{cases} \mathbf{\Sigma}_i^{-1/2}\mathbf{X}_k^i, & 1\le k\le n_1, \\ -\dfrac{n_1}{n_2}\mathbf{\Sigma}_i^{-1/2}\mathbf{Y}_{k-n_1}^i, & n_1+1\le k\le n_1+n_2, \end{cases}$$

and use the same truncation notations as in the proof of Theorem 2.1. With a careful check of the proofs of Theorem 2.1 and Proposition 5.1, we can see that it suffices to show that, for $\mathrm{Card}(\mathbf{N}) = O(x_n^2)$,

$$\mathsf{P}\Big(\max_{1\le i\le m}\|\hat{S}_{ni}^{\{\mathbf{N}\}}\| \ge x_n\sqrt{n_1(1\pm\varepsilon_n x_n^{-2})}\Big) \to \exp\Big(-\frac{1}{\Gamma(d/2)}\exp(-x/2)\Big). \tag{5.12}$$

Let $y_n = x_n\sqrt{n_1(1\pm\varepsilon_n x_n^{-2})}$, where $\varepsilon_n \to 0$ to be specified later. By the Bonferroni inequality, we have for any fixed integer $k$,

$$\begin{aligned}\sum_{l=1}^{2k}(-1)^{l-1}&\sum_{1\le i_1<\cdots<i_l\le m}\mathsf{P}(\|\hat{S}_{ni_1}^{\{\mathbf{N}\}}\|\ge y_n,\dots,\|\hat{S}_{ni_l}^{\{\mathbf{N}\}}\|\ge y_n)\\ &\le \mathsf{P}\Big(\max_{1\le i\le m}\|\hat{S}_{ni}^{\{\mathbf{N}\}}\|\ge y_n\Big)\\ &\le \sum_{l=1}^{2k-1}(-1)^{l-1}\sum_{1\le i_1<\cdots<i_l\le m}\mathsf{P}(\|\hat{S}_{ni_1}^{\{\mathbf{N}\}}\|\ge y_n,\dots,\|\hat{S}_{ni_l}^{\{\mathbf{N}\}}\|\ge y_n).\end{aligned}$$

Theorem 3.1 follows from the following lemma.

LEMMA 5.4. *Let* $\mathrm{Card}(\mathbf{N}) = O(x^2)$. *We have for any fixed* $l$,

$$\begin{aligned}\sum_{1\le i_1<\cdots<i_l\le m}&\mathsf{P}(\|\hat{S}_{ni_1}^{\{\mathbf{N}\}}\|\ge y_n,\dots,\|\hat{S}_{ni_l}^{\{\mathbf{N}\}}\|\ge y_n)\\ &= (1+o(1))\frac{1}{l!}\Big(\frac{1}{\Gamma(d/2)}\exp(-x/2)\Big)^l.\end{aligned}$$

In fact, by Lemma 5.4, we have

$$\begin{aligned}\limsup_{n\to\infty}&\,\mathsf{P}\Big(\max_{1\le i\le m}\|\hat{S}_{ni}^{\{\mathbf{N}\}}\|\ge y_n\Big)\\ &\le 1-\sum_{l=0}^{2k-1}(-1)^l\frac{1}{l!}\Big(\frac{1}{\Gamma(d/2)}\exp(-x/2)\Big)^l\\ &\to 1-\exp\Big(-\frac{1}{\Gamma(d/2)}\exp(-x/2)\Big)\end{aligned}$$

as $k \to \infty$. Similarly,

$$\liminf_{n\to\infty} \mathsf{P}\Big(\max_{1\le i\le m} \|\hat{S}_{ni}^{\{\mathbf{N}\}}\| \ge y_n\Big) \ge 1 - \exp\bigg(-\frac{1}{\Gamma(d/2)}\exp(-x/2)\bigg).$$

This proves Theorem 3.1.

Proof of Lemma 5.4. Let $\mathbf{X}^i = (X_1^i, \ldots, X_d^i)'$ and $\mathbf{Y}^i = (Y_1^i, \ldots, Y_d^i)'$. Put

$$r_{ij} = \max\Big\{\max_{k_1,k_2} |\mathrm{Corr}(X_{k_1}^i, X_{k_2}^j)|, \max_{k_1,k_2} |\mathrm{Corr}(Y_{k_1}^i, Y_{k_2}^j)|\Big\}$$

and

$$\mathcal{I} = \Big\{1 \le i_1 < \cdots < i_l \le m : \max_{1\le k<j\le l} r_{i_k i_j} \ge (\log m)^{-1-\gamma}\Big\}.$$

When $l = 1$, we let $\mathcal{I} = \varnothing$. For $2 \le j \le l-1$, define

$$\begin{aligned}\mathcal{I}_j = \{1 \le i_1 < \cdots < i_l \le m : \mathrm{Card(S)} = j, &\text{ where S is the subset of}\\ \{i_1, \ldots, i_l\} \text{ with the largest cardinality such that } &\forall i_k \ne i_t \in \mathrm{S},\\ &r_{i_k i_t} < (\log m)^{-1-\gamma}\}.\end{aligned}$$

For $j = 1$, define

$$\mathcal{I}_1 = \{1 \le i_1 < \cdots < i_l \le m : r_{i_k i_t} \ge (\log m)^{-1-\gamma} \text{ for every } 1 \le k < t \le l\}.$$

It follows from the definition of $\mathcal{I}_j$ that $\mathcal{I} = \bigcup_{j=1}^{l-1} \mathcal{I}_j$. Then, by (C1), we have $\mathrm{Card}(\mathcal{I}_j) = O(m^{j+2d\rho l})$. Define

$$\mathcal{I}^c = \{1 \le i_1 < \cdots < i_l \le m\} \setminus \mathcal{I}.$$

We have $\mathrm{Card}(\mathcal{I}^c) = C_m^l - O(m^{l-1+2d\rho l}) = (1+o(1))C_m^l$. For $(i_1, \ldots, i_l) \in \mathcal{I}^c$,

$$\bigg\|\frac{1}{n_1}\mathsf{Cov}((\hat{S}_{ni_1}^{\{\mathbf{N}\}}, \ldots, \hat{S}_{ni_l}^{\{\mathbf{N}\}})) - I_{dl}\bigg\| \le C(\log m)^{-1-\gamma} + C(\log m/n)^{(1+\delta)\beta/2}.$$

By Lemma 5.2, the proof of Lemma 5.1 and some tedious calculations,

$$\begin{aligned}&\mathsf{P}(\|\hat{S}_{ni_1}^{\{\mathbf{N}\}}\| \ge y_n, \ldots, \|\hat{S}_{ni_l}^{\{\mathbf{N}\}}\| \ge y_n)\\ &\quad = (1+o(1))\mathsf{P}(\|\mathbf{W}_{i_1}\| \ge y_n/\sqrt{n_1}, \ldots, \|\mathbf{W}_{i_l}\| \ge y_n/\sqrt{n_1}),\end{aligned}$$

where $\mathbf{W}_{i_1}, \ldots, \mathbf{W}_{i_l}$ are independent standard $d$-dimensional random normal vectors. By the tail probabilities of $\chi^2(d)$ distribution,

$$\begin{aligned}&\sum_{\mathcal{I}^c} \mathsf{P}(\|\hat{S}_{ni_1}^{\{\mathbf{N}\}}\| \ge y_n, \ldots, \|\hat{S}_{ni_l}^{\{\mathbf{N}\}}\| \ge y_n)\\ &\quad = (1+o(1))\frac{1}{l!}\bigg(\frac{1}{\Gamma(d/2)}\exp(-y/2)\bigg)^l.\end{aligned} \tag{5.13}$$

To prove the lemma, it suffices to show that for $1 \le j \le l-1$,

$$\sum_{\mathcal{I}_j} \mathsf{P}(\|\hat{S}^{\{\mathbf{N}\}}_{ni_1}\| \ge y_n, \dots, \|\hat{S}^{\{\mathbf{N}\}}_{ni_l}\| \ge y_n) = o(1). \tag{5.14}$$

To keep notation brief, we assume $\mathrm{S} = \{i_{l-j+1}, \dots, i_l\}$ for $(i_1, \dots, i_l) \in \mathcal{I}_j$. Divide $\mathcal{I}_j$ into $\mathcal{I}_{j1}$ and $\mathcal{I}_{j2}$, where

$$\begin{aligned}\mathcal{I}_{j1} = \Bigg\{ & 1 \le i_1 < \cdots < i_l \le m\text{: there exists an } k \in \{i_1, \dots, i_{l-j}\} \\ & \text{such that for some } j_1, j_2 \in \mathrm{S} \text{ with } j_1 \ne j_2,\ r_{kj_1} \ge \frac{1}{(\log m)^{1+\gamma}} \\ & \text{and } r_{kj_2} \ge \frac{1}{(\log m)^{1+\gamma}} \Bigg\}\end{aligned}$$

and $\mathcal{I}_{j2} = \mathcal{I}_j \setminus \mathcal{I}_{j1}$. Then $\mathrm{Card}(\mathcal{I}_{j1}) = O(m^{j-1+4d\rho l})$ and again by Lemma 5.2 and the proof of Lemma 5.1,

$$\begin{aligned}&\sum_{\mathcal{I}_{j1}} \mathsf{P}(\|\hat{S}^{\{\mathbf{N}\}}_{ni_1}\| \ge y_n, \dots, \|\hat{S}^{\{\mathbf{N}\}}_{ni_l}\| \ge y_n) \\ &\quad \le \sum_{\mathcal{I}_{j1}} \mathsf{P}(\|\hat{S}^{\{\mathbf{N}\}}_{ni_{l-j+1}}\| \ge y_n, \dots, \|\hat{S}^{\{\mathbf{N}\}}_{ni_l}\| \ge y_n) \\ &\quad = (1+o(1)) \sum_{\mathcal{I}_{j1}} \mathsf{P}(\|\mathbf{W}_{i_{l-j+1}}\| \ge y_n/\sqrt{n_1}, \dots, \|\mathbf{W}_{i_l}\| \ge y_n/\sqrt{n_1}) \\ &\quad = O(m^{-1+4d\rho l}).\end{aligned}$$

For $(i_1, \dots, i_l) \in \mathcal{I}_{j2}$ and $i_{l-j}$, there is only one $j_1 \in \mathrm{S}$ such that $r_{i_{l-j}j_1} \ge (\log m)^{-1-\gamma}$. For notation briefness, we can assume $j_1 = i_{l-j+1}$. Thus, for any $0 < \varepsilon < 1$, by Theorem 1 in Zaĭtsev (1987),

$$\begin{aligned}&\mathsf{P}(\|\hat{S}^{\{\mathbf{N}\}}_{ni_{l-j}}\| \ge y_n, \dots, \|\hat{S}^{\{\mathbf{N}\}}_{ni_l}\| \ge y_n) \\ &\quad \le \mathsf{P}(\|\tilde{\mathbf{W}}_{i_{l-j}}\| \ge (1-\varepsilon)y_n/\sqrt{n_1}, \dots, \|\tilde{\mathbf{W}}_{i_l}\| \ge (1-\varepsilon)y_n/\sqrt{n_1}) \\ &\qquad + c_1 \exp(-c_2(\log m)^{1+(1-\beta)/2}),\end{aligned} \tag{5.15}$$

where $c_1$ and $c_2$ only depend on $d$ and $\varepsilon$, $(\tilde{\mathbf{W}}_{i_{l-j}}, \dots, \tilde{\mathbf{W}}_{i_l})$ are multivariate norm vector with covariance matrix $\mathsf{Cov}(\hat{S}^{\{\mathbf{N}\}}_{ni_{l-j}}, \dots, \hat{S}^{\{\mathbf{N}\}}_{ni_l})$. By the definition of $\mathcal{I}_{j2}$, we can prove that

$$\begin{aligned}&\left\| \frac{1}{n_1} \mathsf{Cov}(\hat{S}^{\{\mathbf{N}\}}_{ni_{l-j}}, \dots, \hat{S}^{\{\mathbf{N}\}}_{ni_l}) - \begin{pmatrix} \mathbf{D} & 0 \\ 0 & \mathbf{I} \end{pmatrix} \right\| \\ &\qquad \le \frac{C}{(\log m)^{1+\gamma}} + C\left(\frac{\log m}{n}\right)^{(1+\delta)\beta/2},\end{aligned}$$

where $\mathbf{D} = n_1^{-1}\sum_{k=1}^{n_1+n_2} \mathsf{Cov}((\mathbf{Z}_k^{i_{l-j}}, \mathbf{Z}_k^{i_{l-j+1}}))$ and $\mathbf{I}$ is $(j-1)d$-dimensional identity matrix. It follows that

$$\begin{aligned}
&\sum_{\mathcal{I}_{j2}} \mathsf{P}(\|\hat{S}_{ni_1}^{\{\mathbf{N}\}}\| \geq y_n, \ldots, \|\hat{S}_{ni_l}^{\{\mathbf{N}\}}\| \geq y_n) \\
&\qquad \leq \sum_{\mathcal{I}_{j2}} \mathsf{P}(\|\hat{S}_{ni_{l-j}}^{\{\mathbf{N}\}}\| \geq y_n, \ldots, \|\hat{S}_{ni_l}^{\{\mathbf{N}\}}\| \geq y_n) \\
&\qquad \leq (1+o(1)) \sum_{\mathcal{I}_{j2}} m^{-j+1} \mathsf{P}(\|(\tilde{\mathbf{W}}_{i_{l-j}}, \tilde{\mathbf{W}}_{i_{l-j+1}})\| \geq (1-\varepsilon)\sqrt{2}y_n/\sqrt{n_1}) \\
&\qquad\qquad + o(1).
\end{aligned}$$

Since $\max_{1<i<j\leq p}\|\Gamma_{ij}\| \leq r$, we have $\|\mathbf{D}\| \leq 1+r$. This yields that

$$\begin{aligned}
&\mathsf{P}(\|(\tilde{\mathbf{W}}_{i_{l-j}}, \tilde{\mathbf{W}}_{i_{l-j+1}})\| \geq (1-\varepsilon)\sqrt{2}y_n/\sqrt{n_1}) \\
&\qquad \leq C(\log m)^{d/2-1} m^{-2(1-\varepsilon)^2/(1+r)}.
\end{aligned} \tag{5.16}$$

Since $\rho$ is arbitrarily small, we can let $\varepsilon$ satisfy $2(1-\varepsilon)^2/(1+r) > 1+\rho l$. This proves that

$$\sum_{\mathcal{I}_{j2}} \mathsf{P}(\|\hat{S}_{ni_1}^{\{\mathbf{N}\}}\| \geq y_n, \ldots, \|\hat{S}_{ni_l}^{\{\mathbf{N}\}}\| \geq y_n) = O(m^{j+\rho l-j+1-2(1-\varepsilon)^2/(1+r)}) = o(1).$$

Lemma 5.4 is proved. □

5.4. *Proof of Theorem 3.3.* The proof of Theorem 3.3 is given in the supplement material [Liu and Shao (2013)].

5.5. *Proof of Theorem 3.5.* Let $i_0$ be the index such that

$$\|\mathbf{\Sigma}_{i_0}^{-1/2}(\boldsymbol{\mu}_{1i_0} - \boldsymbol{\mu}_{2i_0})\| = \max_{1\leq i\leq m}\|\mathbf{\Sigma}_i^{-1/2}(\boldsymbol{\mu}_{1i} - \boldsymbol{\mu}_{2i})\| \geq \sqrt{(2+\epsilon)\frac{\log m}{n_1}}.$$

Take $\|\theta\| = 1$ such that $\theta'\mathbf{\Sigma}_{i_0}^{-1/2}(\boldsymbol{\mu}_{1i_0} - \boldsymbol{\mu}_{2i_0}) = \|\mathbf{\Sigma}_{i_0}^{-1/2}(\boldsymbol{\mu}_{1i_0} - \boldsymbol{\mu}_{2i_0})\|$. Note that $y_n(\alpha) = 2\log m + (d-2)\log\log m + q_\alpha + o(1)$. We have for any $0 < \varepsilon < \sqrt{1+\epsilon/2} - 1$,

$$\begin{aligned}
\mathsf{P}(\Phi_\alpha^* = 1) &\geq \mathsf{P}(T_{ni_0}^2 \geq y_n(\alpha)) \\
&\geq \mathsf{P}\left(\sum_{k=1}^n \theta'\mathbf{Z}_k^{i_0} \geq (1+\varepsilon)\sqrt{y_n(\alpha)n_1}\right) + o(1) \\
&\geq \mathsf{P}\left(\sum_{k=1}^n \theta'(\mathbf{Z}_k^{i_0} - \mathsf{E}\mathbf{Z}_k^{i_0}) \geq (1+\varepsilon)\sqrt{y_n(\alpha)n_1} - \sqrt{(2+\epsilon)n_1\log p}\right) \\
&\qquad + o(1) \\
&\to 1.
\end{aligned}$$

**Acknowledgements.** We would like to thank the Associate Editor and referees for their insightful comments that have led to significant improvement on the presentation of the paper.

## SUPPLEMENTARY MATERIAL

**Supplement to "A Cramér moderate deviation theorem for Hotelling's $T^2$-statistic with applications to global tests"** (DOI: 10.1214/12-AOS1082SUPP; .pdf). The supplement material includes the moderate deviation result by Sakhanenko (1991), the proof of Theorem 3.3 and the simulation results in Section 4.

## REFERENCES

ANDERSON, T. W. (2003). *An Introduction to Multivariate Statistical Analysis*, 3rd ed. Wiley, Hoboken, NJ. MR1990662

ARRATIA, R., GOLDSTEIN, L. and GORDON, L. (1989). Two moments suffice for Poisson approximations: The Chen–Stein method. *Ann. Probab.* **17** 9–25. MR0972770

CAI, T., LIU, W. and XIA, Y. (2012). Two-sample test of high dimensional means under dependency. Technical report.

CAO, J. and WORSLEY, K. J. (1999). The detection of local shape changes via the geometry of Hotelling's $T^2$ fields. *Ann. Statist.* **27** 925–942. MR1724036

CHEN, S. X. and QIN, Y.-L. (2010). A two-sample test for high-dimensional data with applications to gene-set testing. *Ann. Statist.* **38** 808–835. MR2604697

DE LA PEÑA, V. H., LAI, T. L. and SHAO, Q.-M. (2009). *Self-Normalized Processes: Limit Theory and Statistical Applications*. Springer, Berlin. MR2488094

DEMBO, A. and SHAO, Q.-M. (2006). Large and moderate deviations for Hotelling's $T^2$-statistic. *Electron. Commun. Probab.* **11** 149–159 (electronic). MR2240708

DONOHO, D. and JIN, J. (2004). Higher criticism for detecting sparse heterogeneous mixtures. *Ann. Statist.* **32** 962–994. MR2065195

FAN, J., HALL, P. and YAO, Q. (2007). To how many simultaneous hypothesis tests can normal, Student's $t$ or bootstrap calibration be applied? *J. Amer. Statist. Assoc.* **102** 1282–1288. MR2372536

GERARDINA, E., CHÉTELATD, G., CHUPIN, M., CUINGNET, R., DESGRANGES, B., KIME, H. S., NIETHAMMER, M., DUBOIS, B., STÉPHANE LEHÉRICY, S., LINE GARNERO, L., EUSTACHE, F. and COLLIOT, O. (2009). Multidimensional classification of hippocampal shape features discriminates Alzheimer's disease and mild cognitive impairment from normal aging. *NeuroImage* **47** 1476–1486.

HALL, P. and JIN, J. (2010). Innovated higher criticism for detecting sparse signals in correlated noise. *Ann. Statist.* **38** 1686–1732. MR2662357

HALL, P. and WANG, Q. (2010). Strong approximations of level exceedences related to multiple hypothesis testing. *Bernoulli* **16** 418–434. MR2668908

JING, B.-Y., SHAO, Q.-M. and WANG, Q. (2003). Self-normalized Cramér-type large deviations for independent random variables. *Ann. Probab.* **31** 2167–2215. MR2016616

LIN, Z. and LIU, W. (2009). On maxima of periodograms of stationary processes. *Ann. Statist.* **37** 2676–2695. MR2541443

LIU, W.-D., LIN, Z. and SHAO, Q.-M. (2008). The asymptotic distribution and Berry–Esseen bound of a new test for independence in high dimension with an application to stochastic optimization. *Ann. Appl. Probab.* **18** 2337–2366. MR2474539

Liu, W. and Shao, Q.-M. (2010). Cramér-type moderate deviation for the maximum of the periodogram with application to simultaneous tests in gene expression time series. *Ann. Statist.* **38** 1913–1935. MR2662363

Liu, W. and Shao, Q. M. (2013). Supplement to "A Cramér moderate deviation theorem for Hotelling's $T^2$-statistic with applications to global tests." DOI:10.1214/12-AOS1082SUPP.

Martens, J. W. M., Nimmrich, I., Koenig, T., Look, M. P., Harbeck, N., Model, F., Kluth, A., de Vries, J. B., Sieuwerts, A. M., Portengen, H., Meijer-Van Gelder, M. E., Piepenbrock, C., Olek, A., Höfler, H., Kiechle, M., Klijn, J. G. K., Schmitt, M., Maier, S. and Foekens, J. A. (2005). Association of DNA methylation of Phosphoserine Aminotransferase with response to endocrine therapy in patients with recurrent breast cancer. *Cancer Research* **65** 4101–4117.

Sakhanenko, A. I. (1991). Berry–Esseen type estimates for large deviation probabilities. *Sib. Math. J.* **32** 647–656.

Shao, Q.-M. (1999). A Cramér type large deviation result for Student's $t$-statistic. *J. Theoret. Probab.* **12** 385–398. MR1684750

Storey, J. D. and Tibshirani, R. (2001). Estimating false discovery rates under dependence, with applications to DNA microarrays. Technical report.

Styner, M., Oguz, I., Xu, S., Brechbühler, C., Pantazis, D., Levitt, J. J., Shenton, M. E. and Gerig, G. (2006). Framework for the statistical shape analysis of brain structures using SPHARM-PDM. *Insight Journal* **1071** 242–250.

Taylor, J. E. and Worsley, K. J. (2008). Random fields of multivariate test statistics, with applications to shape analysis. *Ann. Statist.* **36** 1–27. MR2387962

Zaĭtsev, A. Y. (1987). On the Gaussian approximation of convolutions under multidimensional analogues of S. N. Bernstein's inequality conditions. *Probab. Theory Related Fields* **74** 535–566. MR0876255

Zhao, Z., Taylor, W. D., Styner, M., Steffens, D. C., Krishnan, K. R. R. and MacFall, J. R. (2008). Hippocampus shape analysis and late-life depression. *PLoS One* **3** e1837.

Department of Mathematics
Institute of Natural Sciences
Shanghai Jiao Tong University
Shanghai, P.R. China
E-mail: liuweidong99@gmail.com

Department of Statistics
Chinese University of Hong Kong
Shatin, N.T.
Hong Kong, P.R. China
E-mail: qmshao@cuhk.edu.hk